United States Environmental Protection Agency

Office Of Air Quality Planning And Standards Research Triangle Park, NC 27711

EPA-454/R-99-004
May 1999

Air

EPA

DRAFT GUIDANCE ON THE USE OF MODELS AND OTHER ANALYSES IN ATTAINMENT DEMONSTRATIONS FOR THE 8-HOUR OZONE NAAQS

EPA-454/R-99-004

DRAFT GUIDANCE ON THE USE OF MODELS AND OTHER ANALYSES IN ATTAINMENT DEMONSTRATIONS FOR THE 8-HOUR OZONE NAAQS

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Radiation
Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711

MAY 1999

DISCLAIMER

This document includes draft recommendations for using models and other analyses to help identify strategies which are effective for meeting an 8-hour national ambient air quality standard promulgated in 1997. Pending outcome of litigation regarding these NAAQS, recommendations contained herein may not be enforced and are subject to change. Readers should review the "Foreword" section to the document to understand the context within which this information is being presented. Mention of trade names or commercial products in this document is not intended to constitute endorsement or recommendation for use.

ACKNOWLEDGMENTS

We would like to acknowledge contributions from members of an external review group, the STAPPA/ALAPCO emissions/modeling committee and U.S. EPA Regional Office modeling staffs in providing detailed comments and suggestions regarding the draft version of this guidance. In particular, we would like to thank staff members of the Lake Michigan Air Directors Consortium (LADCO), Microelectronics Center of North Carolina (MCNC), South Coast Air Quality Management District (SCAQMD), California Air Resources Board (CARB), Texas Natural Resources Conservation Commission (TNRCC) and Dynetl Corporation for testing our ideas for a modeled attainment test and sharing the results with us.

Table of Contents

DISCLAIMER iii

ACKNOWLEDGMENTS iv

List Of Figures ix

List Of Tables x

Foreword xi

Executive Summary 1

1.0 Introduction 19
1.1 What Is The Purpose Of This Document? 19
1.2 Does The Guidance In This Document Apply To Me? 19
1.3 How Does The Perceived Nature Of Ozone Affect My Attainment Demonstration? 20
1.4 What Topics Are Covered In This Guidance? 22

2.0 What Is A Modeled Attainment Demonstration?--An Overview 27
2.1 What Is The Recommended Modeled Attainment Test?--An Overview 27
2.2 What Does A Recommended Weight Of Evidence Determination Consist Of?--An Overview 28
2.3 Why Should A Model Be Used In A "Relative" Sense And Why May Corroboratory Analyses Be Used In A Weight Of Evidence Determination? 29

3.0 What Is The Recommended Modeled Attainment Test? 33
3.1 How Do I Select Appropriate Inputs For The Modeled Attainment Test? 34
3.2 How Do I Apply The Recommended Modeled Attainment Test? 44
3.3 What Are Key Implications Of The Recommended Modeled Attainment Test? 47
3.4 What Is A Screening Test, And Why Is It Needed? 49

4.0 If I Use A Weight Of Evidence Determination, What Does This Entail? 51
4.1 What Analyses Should I Consider In A Weight Of Evidence Determination? 52
4.1.1 Air Quality Models 52
4.1.2 Analysis Of Air Quality And Emissions Trends 57
4.1.3 Use Of Observational Models 58
4.2 What If I Want To Consider Additional Corroborative Analyses? 60

5.0 How Can I Improve Modeling And Other Analyses In Weight Of Evidence Determinations? 63

5.1 What Data Gathering Or Other Efforts Might Be Helpful To Support Current Analyses Or Subsequent Reviews? 63
5.2 Why Is It Desirable To Plan For A Subsequent Review? 66
5.2.1 Integration With Attainment Strategies For 1-Hr Ozone NAAQS 66
5.2.2 Anticipated Modeling Principles For $PM_{2.5}$ And Visibility, And Integrating Ozone Strategies With Goals For $PM_{2.5}$ And Visibility 66

6.0 What Documentation Do I Need To Support My Attainment Demonstration? 69

7.0 References Cited In Part I And In Section 1.0 75

8.0 How Do I Apply Air Quality Models?-- An Overview 81

9.0 How Do I Get Started? 87
9.1 What Is A "Conceptual Description"? 87
9.2 What Sorts Of Analyses Might Be Useful For Developing And Refining A Conceptual Description? 90

10.0 What Does A Modeling/Analysis Protocol Do, And What Does Developing One Entail? 93
10.1 What Is The Protocol's Function? 93
10.2 What Subjects Should Be Addressed In The Protocol? 93

11.0 What Should I Consider In Choosing An Air Quality Model? 95
11.1 What Prerequisites Should An Air Quality Model Meet To Qualify For Use In An Attainment Demonstration? 95
11.2 What Factors Affect My Choice of A Model For A Specific Application? 97
11.3 What Are Some Examples Of Air Quality Models Which May Be Considered? .. 99

12.0 How Do I Decide Which Meteorological Episodes To Model? 101
12.1 What Are The Most Important Criteria For Choosing Episodes? 101
12.2 What Additional, Secondary Criteria May Be Useful For Selecting Episodes? .. 104

13.0 What Should I Consider When Selecting A Modeling Domain And Its Horizontal/Vertical Resolution? 107
13.1 How Do I Choose Between An Urban Scale Or Regional Domain? 107
13.2 What Horizontal Grid Cell Size Is Necessary? 108
13.3 How Many Vertical Layers Should I Consider? 110
13.4 What Else Should I Consider In Choosing Finely And Coarsely Resolved Portions Of Nested Regional Models? 111

14.0 How Do I Produce Meteorological and Air Quality Inputs Needed By An Air Quality Model? 113

14.1 What Approaches Are Available For Generating Meteorological Data? 113
14.2 How Do I Deal With Data Management And Computer-related Constraints When Applying Dynamic Meteorological Models? 115
14.3 How Do I Quality Assure Results Generated By A Meteorological Model? 116
14.4 What Are Some Past Applications Of Dynamic Meteorological Models? 118
14.5 How Do I Address An Air Quality Model's Need For Air Quality Inputs? 118

15.0 How Do I Produce Emission Inputs Needed For An Air Quality Model? 121
15.1 What Countywide Emission Estimates Are Needed To Support Air Quality Models? 122
15.2 Can I Use the National Emissions Trends Inventory As A Starting Point? 123
15.3 How Do I Convert Countywide Inventory Information Into Data Used In Air Quality Models? 123
15.4 What Should I Do To Quality Assure Emissions Estimates? 126
15.5 How Do I Estimate Emissions For Future Years? 127

16.0 How Do I Assess Model Performance And Make Use Of Diagnostic Analyses? 129
16.1 How Can I Evaluate Performance Of An Air Quality Model? 129
16.1.1 How Can I Use Graphics To Make A "Big Picture" Assessment Of Model Performance? 130
16.1.2 How Can Ozone Metrics Be Used To Assess Model Performance? ... 131
16.1.3 How Can I Use Available Precursor Observations To Evaluate Model Performance? 133
16.1.4 How Can I Use Corroborative Analyses With Observational Models To Help Evaluate Air Quality Model Performance? 133
16.1.5 What Data Bases Reflecting Changes In Emissions Are Available To Evaluate Model Performance? 134
16.1.6 How Do I Use Ratios of Indicator Species To Evaluate Model Performance? 135
16.1.7 Are Retrospective Analyses Useful For Evaluating Model Performance? 136
16.1.8 All Of These Performance Tests Have Shortcomings, So What Do I Do? 136
16.2 How Can I Make Good Use of Diagnostic Tests? 141
16.2.1 Use Of Sensitivity Tests 141
16.2.2 Use of Process Analysis 142

17.0 References Cited In Part II 147

Glossary 155

List Of Figures

Figure 3.1. Relationship Between Grid Resolution And Grid Cells Considered To Be In The Vicinity Of A Monitor----------38

Figure 3.2. Choosing Predictions To Estimate RRF's----------40

Figure 3.3. Mean Relative Reduction As A Function Of Mean Predicted Current 8-Hour Daily Maxima----------42

Figure 4.1. Examples Showing Use Of Trend Analysis In Attainment Demonstrations---------59

List Of Tables

Table ES.1. Recommended Core Analyses For A Weight Of Evidence Determination---------5

Table ES.2. Recommended Documentation For Demonstrating Attainment Of The 8-Hour Ozone NAAQS----------8

Table 2.1. Guidelines For Using A Weight Of Evidence Determination----------28

Table 3.1. Example Monitoring Data For Nonattainment Area "A"----------34

Table 3.2. Example Illustrating Selection Of Current Design Values----------36

Table 3.3. Default Recommendations For Nearby Grid Cells Used To Calculate RRFs-------37

Table 3.4. Example Calculation Of A Site-Specific Future Design Value $(DVF)_I$----------46

Table 4.1. Recommended Core Analyses For A Weight Of Evidence Determination, Factors Affecting Their Credibility And Outcomes Consistent With Meeting The NAAQS----------53

Table 6.1. Recommended Documentation For Demonstrating Attainment Of The 8-hour NAAQS For Ozone----------70

Table 11.1. Some Air Quality Models Used To Model Ozone----------100

Table 14.1. Some Past Applications Of Dynamic Meteorological Models----------119

Table 15.1. Some Emissions Models And Example Applications----------125

Table 16.1. Summary Of Methods To Evaluate Performance Of Air Quality Models----------137

Table 16.2. Potentially Useful Diagnostic Tests At Various Stages Of Modeling----------144

Foreword

Readers should note that the U.S. Court of Appeals for the District of Columbia Circuit issued a decision on May 14, 1999 which prohibited enforcement of the 8-hr ozone national ambient air quality standard (NAAQS). Therefore, this draft guidance is being distributed at this time to document the position of the Agency when the Court Decision was issued. Moreover, the concepts developed in the guidance are applicable to other multi-hour standards. When the litigation issues have been addressed and resolved, we will review this draft guidance and revise it as appropriate. Distribution is being made since this document is the product of an extensive dialogue with the stakeholders from the modeling community and since we have had numerous requests for release of the current version. Stakeholders involved in the development of the draft include representatives from State and local governmental agencies, academia, consultants, industry, and environmental organizations in addition to the EPA Regional Offices.

Distribution of this document should not be construed as implementation of the 8-hr standard; it is merely provided for information to interested parties and to document the status of the guidance at the time of the court decision. Thus, we feel it is appropriate to provide this draft guidance document to State and local agencies as well as other requestors. However, no actions should be taken to implement this draft guidance until the litigation issues have been resolved and final guidance has been issued.

Executive Summary

Readers should note that the U.S. Court of Appeals for the District of Columbia Circuit issued a decision on May 14, 1999 which prohibited enforcement of the 8-hr ozone national ambient air quality standard (NAAQS). Therefore, this draft guidance is being distributed at this time to document the position of the Agency when the Court Decision was issued. Moreover, the concepts developed in the guidance are applicable to other multi-hour standards. When the litigation issues have been addressed and resolved, we will review this draft guidance and revise it as appropriate. Distribution is being made since this document is the product of an extensive dialogue with the stakeholders from the modeling community and since we have had numerous requests for release of the current version. Stakeholders involved in the development of the draft include representatives from State and local governmental agencies, academia, consultants, industry, and environmental organizations in addition to the EPA Regional Offices.

Distribution of this document should not be construed as implementation of the 8-hr standard; it is merely provided for information to interested parties and to document the status of the guidance at the time of the court decision. Thus, we feel it is appropriate to provide this draft guidance document to State and local agencies as well as other requestors. However, no actions should be taken to implement this draft guidance until the litigation issues have been resolved and final guidance has been issued.

This document recommends procedures for estimating if a control strategy to reduce emissions of ozone precursors will lead to attainment of the 8-hour national ambient air quality standard (NAAQS) for ozone. The document also describes how to apply air quality models to generate the predictions later used to see if attainment is shown. Guidance in this document applies to "traditional" nonattainment areas, as well as to "transitional" nonattainment areas for which modeling is needed or desired. This includes locations that violate the 8-hour ozone NAAQS and may or may not violate the 1-hour NAAQS.

The guidance consists of two major parts. Part I addresses the question, "how do I use the results of models and other analyses to help demonstrate attainment?" It explains what we mean by a *modeled attainment demonstration*, a *modeled attainment test*, a *screening test* and a *weight of evidence determination*. It also identifies additional data which, if available, should enhance credibility of model results and results of other analyses used in a weight of evidence determination. Part I concludes by identifying what States should include in documentation accompanying an attainment demonstration. Sections ES1.0 - ES6.0 summarize the contents of Part I of the guidance.

Part II of the guidance describes how to apply air quality models. The output from such a model is then used to support an attainment demonstration, as described in Part I of the guidance. The recommended procedure for applying a model has 8 steps.

1. Develop a conceptual description of the problem to be addressed.

2. Develop a modeling/analysis protocol.
3. Select an appropriate model to support the demonstration.
4. Select appropriate meteorological episodes to model.
5. Choose an appropriate area to model with appropriate horizontal/vertical resolution.
6. Generate meteorological and air quality inputs to the air quality model.
7. Generate emissions inputs to the air quality model.
8. Evaluate performance of the air quality model and perform diagnostic tests.

After these steps are completed, the model is used to simulate effects of candidate control strategies. Model applications require a substantial effort. States should work closely with the appropriate U.S. EPA Regional Office(s) in executing each step. This will increase the likelihood of approval of the demonstration at the end of the process. Sections ES7.0 - ES14.0 summarize Part II of the guidance.

ES 1.0. What Is An Attainment Demonstration?

An attainment demonstration consists of (a) analyses which estimate whether selected emissions reductions will result in ambient concentrations that meet the NAAQS, and (b) an identified set of measures which will result in the required emissions reductions. This guidance describes how to use air quality models and other analyses to determine if results of a simulated control strategy indicate attainment. Determining necessary emission reductions may be done by relying exclusively on results obtained with air quality models. These include the outcomes of a modeled attainment test plus a screening test to estimate whether a proposed emission reduction suffices to meet the NAAQS. Other analyses, including trends analyses, observational models, etc. may be used to supplement the modeled attainment and screening tests.

A modeled attainment test compares ozone concentration predictions with the ozone NAAQS. The NAAQS is met if the 4th highest 8-hour daily maximum ozone concentration, averaged over 3 consecutive years, is $\leq$ 0.08 ppm. The average 4th highest 8-hour daily maximum concentration is called the "design value" for ozone. The modeled attainment test is passed if predicted future design values near all monitoring sites are $\leq$ 84 ppb. A screening test, applied at other locations with consistently high model predictions, may also be needed.

Provided the modeled attainment test and supplementary screening test are passed or close to being passed, States may use a broader set of analyses to estimate if attainment is likely. This is called a "weight of evidence determination". A weight of evidence determination combines results of the modeled attainment and screening tests with other results obtained with air quality models, as well as conclusions drawn from analyzing monitored air quality data, emissions estimates and meteorological data. Results of each analysis are considered in concert to determine whether or not attainment is likely.

ES 2.0. What Is The Recommended Modeled Attainment Test?

The recommended modeled attainment test uses monitored design values in concert with model-generated data. The test uses model results in a "relative" rather than "absolute" sense, and is applied near ozone monitoring sites. Eight (8)-hour daily maximum ozone is predicted near a monitor for each day in the test. Each day's prediction is then summed and averaged, first with current emissions and then with future emissions. The ratio of future to current predicted mean 8-hour daily maxima is then computed. This ratio is called the "relative reduction factor", or RRF. The test consists of 4 steps.

1. Compute a current site-specific design value from *monitored* data.

2. Use air quality modeling results to estimate a site-specific relative reduction factor.

3. Multiply the relative reduction factor obtained in step 2 times the site-specific design value in step 1. The result is a predicted site-specific future design value. If this value is ≤ 84 ppb, the test is passed at the monitor site being evaluated.

4. Repeat steps 1 through 3 for each monitoring site where the current monitored design value approaches or exceeds 84 ppb. If predicted future site-specific design values are ≤ 84 ppb at each site, the test is passed.

The modeled attainment test is described in Section 3.0.

ES3.0 What Is The Screening Test, And Why Is It Needed?

The modeled attainment test does not address future air quality at locations where there is no nearby ozone monitor. If the air quality model *consistently* predicts 8-hour daily maximum ozone concentrations at a particular, unmonitored location which are substantially higher than any predicted near a monitoring site, a State should perform an additional screening test. The screening test is to multiply the area wide monitored design value (i.e., the highest of the site-specific monitored design values) times the relative reduction factor(s) predicted at the suspect receptor location(s). If the resulting estimated future design value(s) is ≤ 84 ppb, the outcome is not inconsistent with attainment of the NAAQS. The screening test is discussed in Section 3.4.

ES 4.0. What Is A Weight Of Evidence Determination?

A State should always utilize available air quality, meteorological and emissions data to complement a modeling analysis. These data are used to develop a conceptual description of an area's nonattainment problem. This description is useful for guiding a modeling analysis. If the modeled attainment and screening tests are passed or nearly passed a State may choose to use a weight of evidence (WOE) determination to estimate if attainment is likely.

A weight of evidence determination is a diverse set of technical analyses performed to corroborate findings of the modeled attainment and screening tests. If a weight of evidence determination is used, a State should consider a recommended core set of analyses consisting of (1) a set of air quality model results which includes the previously described tests plus additional analyses of estimated concentrations, (2) an analysis of observed air quality and estimated emissions trends, and (3) an analysis of outcomes produced by observational models.

We identify factors which enhance credibility of evidence produced by each of the core analyses, as well as outcomes which would be consistent with the likelihood that a strategy demonstrates attainment. This is illustrated in Table ES.1 for the 3 recommended core analyses.

A State may include other types of analyses, in addition to the core analyses, in a weight of evidence determination. For another analysis to be considered three criteria should be satisfied:

(1) a State should discuss why the proposed analysis is relevant to assessing attainment,

(2) a State should identify the procedure to be used and the data base available to support it, and

(3) a State should identify (in advance) outcomes which would be consistent with a hypothesis that a proposed strategy will lead to attainment.

Weight of evidence and its use is discussed in Section 4.0.

ES 5.0. Why Do We Recommend This Modeled Attainment Test And Offer An Option To Perform A Weight Of Evidence Determination?

Test results can be readily related to the form of the NAAQS. The NAAQS is concerned with the 4th highest 8-hour daily maximum concentration averaged over 3 consecutive years. It is difficult to tell whether a modeled exceedance in a particular episode is or is not inconsistent with meeting the NAAQS. Thus, using a model by itself to rigorously assess whether the NAAQS is met would require modeling a substantial number of days in three different years. Further, we believe relatively resource-intensive models are needed to simulate effects of reducing precursor emissions on ozone. Thus, the test uses *observed* design values to "anchor" model predictions to the form of the NAAQS. Design values are, by definition, calculated consistently with the form of the NAAQS, and their use allows a State to apply a resource-intensive model to see how they might be changed by a control strategy.

The approach recognizes uncertainty in model predictions. Problems in interpreting model results posed by uncertainty in the predictions may be greater for addressing the 8-hour NAAQS than was true for the 1-hour standard. The 8-hour NAAQS is closer to continental background values. Further, design values tend to be closer to the specified level of the NAAQS than is true for the 1-hour standard. As a result, the signal (i.e., the change we wish to effect in

Table ES.1. Recommended Core Analyses For A Weight Of Evidence Determination

(1) Type Of Analysis	(2) Factors Increasing Credibility Of The Analysis	(3) Outcomes Consistent With Hypothesis That A Candidate Strategy Will Lead To Attainment
Air Quality Models	-good model performance -extensive observational data base available -short projection periods -carefully quality assured inventory -confidence in meteorological inputs -good ability to pose and address questions about a strategy's adequacy -other analyses tend to corroborate conclusions -selected episode days have observations near the design value	-the modeled attainment test is passed -screening estimates for future design values at sites w/o monitors are ≤84 ppb -the attainment/screening tests are nearly passed, the control strategy requires additional reductions and efforts are underway to subsequently review/refine the strategy -commitment is made to deploy monitors at locations not passing the screening test -substantial modeled improvement in air quality is predicted using several measures described in Section 4.1.1. -similar conclusions are reached with other peer reviewed models

Table ES.1. Recommended Core Analyses For A Weight Of Evidence Determination (concluded)

(1) Type Of Analysis	(2) Factors Increasing Credibility Of The Analysis	(3) Outcomes Consistent With Hypothesis That A Candidate Strategy Will Lead To Attainment
Analysis of Air Quality and Emissions Trends	-current or future (air quality model) predicted design value is within a few ppb above 84 ppb -extensive monitoring network exists -both ozone and precursor trends are available -statistical model used to normalize trend for meteorological differences explains much variance -short projection periods are used in the analysis -a pronounced, statistically significant downward trend is apparent -similar conclusions are reached using several trend parameters -continued, comparable relative reductions in emissions are provided for	-a pronounced downward normalized trend exists in the site-specific design value at all sites with design values greater than 84 ppb -Using projected emissions to extrapolate the air quality trend line to the required attainment date indicates an 8-hour daily maximum concentration ≤ 84 ppb. -Other observed air quality trend parameters also show a substantial improvement.
Use of Observational Models	-an extensive monitoring network exists -precursor and indicator species are measured using instruments with appropriate sensitivity -monitoring sites appear spatially representative -data have been quality assured, and results are self-consistent -plausible physical explanations exist for findings	-Findings indicate sources controlled in the candidate strategy are important causes of observed high ozone -Analysis of indicator species suggests the direction of the strategy (e.g., emphasis on VOC or NOx) is appropriate.

the design value) to noise (i.e., uncertainty in predictions) ratio may be smaller than heretofore.

The recommended modeled attainment test reduces uncertainty (i.e., "noise") in three important ways. First, monitored data (i.e., current design values) are incorporated directly into the test. These data are likely measured with greater accuracy than an absolute model prediction, and precision of the measurements is known. Second, the outcome of the test is based on a composite set of calculations from several modeled days rather than a single day. This reduces the risk of choosing an inappropriate strategy on the basis of a single outcome which is subject to uncertainty. Third, if the outcome of the test is close to pass/fail, a weight of evidence determination may be used to see whether other model outputs and other types of analyses provide corroborative evidence for conclusions drawn from the test.

ES 6.0: What Documentation Is Needed To Support An Attainment Demonstration?

A State should address 9 subject areas in its documentation. These are enumerated in Table ES.2. Documentation should be accompanied by an executive summary which addresses each of the 9 areas shown in the table. Documentation requirements are addressed in Section 6.0.

ES7.0. What Is A "Conceptual Description"?

A conceptual description is a qualitative way of characterizing an area's nonattainment problem. For example, is the problem one dominated by local emissions or are regional factors also important? Do sites violating the NAAQS reflect a spatial or temporal pattern in some way?, etc. A conceptual description is based on use of readily available air quality, meteorological and emissions information. It may be refined later as additional analyses are performed. States should develop a conceptual description of each nonattainment problem they are seeking to solve as an initial step in developing a solution. It serves as a means for guiding later decisions which are needed in a modeling analysis. Suggestions for developing a conceptual description are found in Section 9.0.

ES8.0. What Does A Modeling/Analysis Protocol Do And What Should It Contain?

A modeling/analysis protocol is a document which identifies methods and procedures to be used in the analyses. The protocol also identifies ground rules to be followed in undertaking analyses to estimate emission reductions needed to meet the NAAQS. Ground rules include a description of how affected stakeholders in the modeling/analysis process will be encouraged to participate, the process by which decisions will be made, means used for communicating issues and decisions, and the methods, data bases and procedures to be used to obtain results. As the name implies, the protocol should address use of other analyses as well as air quality modeling. The document is usually prepared by the State/local agency(ies) having lead responsibility for the

Table ES.2. Recommended Documentation For Demonstrating Attainment Of The 8-hour Ozone NAAQS

Subject Area	Purpose of Documentation	Issues Included
Modeling/Analysis Protocol	Communicate scope of the analysis and document stakeholder involvement	Names of stakeholders participating in preparing and implementing the protocol; Types of analyses performed; Steps followed in each type of analyses; Days and domain considered.
Emissions Preparations and Results	Assurance of valid, consistent and emissions data base. Appropriate procedures are used to derive emission estimates needed for air quality modeling.	Data base used and quality assurance methods applied; Data processing used to convert data base to model-compatible inputs; Deviations from existing guidance and underlying rationale; VOC, NOx, CO emissions by State/county for major source categories.
Air Quality/Meteorology Preparations and Results	Assurance that representative air quality and meteorological inputs are used in analyses	Extent of data base and procedures used to derive & quality assure inputs for analyses used in the weight of evidence determination; Departures from guidance and their underlying rationale; Performance of the meteorological model, if used to generate meteorological inputs to the air quality model.

Table ES.2. Recommended Documentation For Demonstrating Attainment Of The 8-hour Ozone NAAQS (continued)

Subject Area	Purpose of Documentation	Issues Included
Performance Evaluation for Air Quality Model (and Other Analyses)	Show decision makers and the public how well the model (or other analyses) reproduced observations or otherwise performed on the days selected for analysis	Summary of observational data base available for comparison; Identification of performance tests used and their results; Ability to reproduce observed temporal and spatial patterns; Overall assessment of what the performance evaluation implies.
Diagnostic Tests	Ensure rationale used to adjust model inputs or to discount certain results is physically justified and the remaining results make sense.	Results from application prior to adjustments; Consistency with scientific understanding and expectations; Tests performed, changes made and accompanying justification; Short summary of final predictions.

Table ES.2. Recommended Documentation For Demonstrating Attainment Of The 8-hour Ozone NAAQS (continued)

Subject Area	Purpose of Documentation	Issues Included
Description of the Strategy Demonstrating Attainment	Provide the EPA and the public an overview of the plan selected in the attainment demonstration.	Qualitative description of the attainment strategy; Reductions in VOC, NOx, and/or CO emissions from each major source category for each State/county from current (identify) emission levels; Clean Air Act mandated reductions and other reductions; Show predicted 8-hr site-specific future design values for the selected control scenario and identify any location which fails the screening test described in Section 3.4; Identification of authority for implementing emission reductions in the attainment strategy. Evidence that emissions remain at or below projected levels throughout the 3-year period used to determine future attainment.
Data Access	Enable the EPA or other interested parties to replicate model performance and attainment simulation results, as well as results obtained with other analyses.	Assurance that data files are archived and that provision has been made to maintain them; Technical procedures for accessing input and output files; Identify computer on which files were generated and can be read, as well as software necessary to process model outputs; Identification of contact person, means for downloading files and administrative procedures which need to be satisfied to access the files.

Table ES.2. Recommended Documentation For Demonstrating Attainment Of The 8-hour Ozone NAAQS (concluded)

Subject Area	Purpose of Documentation	Issues Included
Weight of Evidence Determination	Assure the EPA and the public that the strategy meets applicable attainment tests and is likely to produce attainment of the NAAQS within the required time.	Description of the modeled attainment test and observational data base used; Identification of air quality simulation model used; Identification of other analyses performed; Outcome of each analysis, including the modeled attainment test; Assessment of the credibility associated with each type of analysis in this application; Narrative describing process used to conclude the overall weight of available evidence supports a hypothesis that the selected strategy is adequate to attain the NAAQS.
Review Procedures Used	Provide assurance to the EPA and the public that analyses performed in the attainment demonstration reflect sound practice	Scope of technical review performed by those implementing the protocol; Assurance that methods used for analysis were peer reviewed by outside experts; Conclusions reached in the reviews and the response thereto.

modeling/analysis, in consultation with stakeholders. The protocol should be kept up to date to reflect major subsequent decisions made after the initial version is completed. Specific topics which should be included in the protocol are identified in Section 10.0.

ES9.0. What Should I Consider In Choosing An Air Quality Model?

Several prerequisites need to be met for a model to qualify for use in supporting an

attainment demonstration.

1. The model has received a scientific peer review.

2. The model can be demonstrated to be applicable to the problem on a theoretical basis.

3. Data bases needed to perform the analysis are available and adequate.

4. Available past appropriate performance evaluations have shown the model is not biased toward underestimates.

5. A protocol on methods and procedures to be followed has been established.

6. The model is available to users for free or at a reasonable cost, and is not proprietary.

To select a qualifying model for a particular application, States should first determine what attributes are needed for a qualifying model to address the nonattainment area's ozone problem, and then choose among models possessing these attributes. Five factors should be considered in selecting an air quality model for a specific application. These are listed in approximate order of importance.

1. Nature of the air quality problem leading to nonattainment of the ozone NAAQS should first be assessed, and the selected model should have attributes and capabilities consistent with the perceived nature of the problem.

2. Availability, documentation and past performance should be satisfactory.

3. Relevant experience of available staff and contractors should be consistent with choice of a model.

4. Time and resource constraints may be considered.

5. Consistency of the model with what is used in adjacent regional applications should be considered.

Choice of a model should be concurred with by the appropriate U.S. EPA Regional Office and U.S. EPA Model Clearinghouse.

Prior to using model results in a specific attainment demonstration, a State should show that the model performs adequately in replicating base case observations available for that demonstration. Further discussion of model selection occurs in Section 11.0.

ES10.0. How Do I Choose Meteorological Episodes?

States should consider four primary criteria when choosing meteorological episodes for modeling. Tradeoffs among these may often be necessary. Such tradeoffs need to be resolved on a case by case basis.

1. Choose frequently occurring episodes containing days reflecting a variety of wind orientations observed to occur when 8-hour daily maxima exceed 84 ppb at one or more monitors.

2. Choose episodes containing days with observed 8-hour daily maximum ozone concentrations close to (e.g., ± 10 ppb) the average 4th high daily maximum observed at monitoring sites during a 3-year period straddling the period from which each episode is drawn (i.e., days approximately as severe as implied by the form of the NAAQS).

3. Choose episodes containing days for which measurements aloft, measurements of indicator species and/or precursor measurements exist.

4. Choose a sufficient number of days so that several days are available for use in the modeled attainment test for each monitoring site where the NAAQS is violated.

States may be able to resolve conflicts among the primary criteria for selecting episodes by considering one or more secondary criteria. The following are identified as secondary criteria. States may identify, document and present the rationale for criteria in addition to these.

1. Give preference to previously modeled episodes.

2. Give preference to episodes occurring during the period corresponding to the current design value.

3. Give preference to episodes maximizing the number of days and sites observing 8-hour daily maxima close to the level of severity specified in the NAAQS.

4. Include weekends among the selected days, especially if daily maxima exceeding 84 ppb are observed on such days.

5. If applying a regional model, choose episodes meeting the other primary and secondary criteria in as many nonattainment areas as possible.

Episode selection is discussed in Section 12.0.

ES11.0. How Do I Select A Modeling Domain And Its Horizontal/Vertical Resolution?

States should review available air quality, meteorological and emissions data to help select a domain size which is consistent with a nonattainment area's problem but which is not unnecessarily resource intensive. We suggest a procedure for comparing regional (upwind) observations with local design values which may be useful in choosing between regional and urban scale domains. A typical urban domain may be about 300 km on a side. A typical regional domain exceeds 1000 km on a side.

Choice of horizontal/vertical resolution presents a conflict between resources and data base management vs. scientific rigor. Sensitivity of results to grid resolution may vary on a case by case basis. For urban scale analyses and for the fine portion of a nested regional application, grid cells as small as 4 km may be preferable. However, for practical reasons, we allow flexibility in choice of grid cell size, so long as the cells are 4-12 km or smaller. Coarse portions of regional applications may use 36-km grid cells or smaller. If cells as large as 12 km are used for an urban analysis or in a "fine" portion of a nested grid, States should perform a test to assess sensitivity of conclusions drawn simulating a tentatively selected strategy to use of such large cells, if feasible.

Models should ordinarily include at least 7-9 vertical layers within the planetary boundary layer (PBL), and 1-2 layers above the PBL. Care should be taken to place vertical layers so that the maximum afternoon mixing height can be estimated as precisely as feasible. Tests assessing sensitivity of conclusions drawn with a tentatively selected strategy to use of more vertical layers should be performed, if feasible.

More detailed suggestions for selecting domain size and grid resolution are contained in Section 13.0.

ES12.0. How Do I Produce Meteorological And Air Quality Inputs Required By An Air Quality Model?

We recommend that States use a dynamic meteorological model with four dimensional data assimilation (FDDA) as the principal means for generating meteorological inputs required by air quality models used in ozone attainment demonstrations. Any such meteorological model which has received a scientific peer review may be used. As with the output from emissions models, it is critical that results of meteorological models be quality assured. We identify several potentially useful means for doing so:

1. comparison with upper air measurements "held back" from use in FDDA;

2. comparison of calculated trajectories with observed air quality patterns;

3. use of computer graphics to discern spatial discrepancies;

4. simulation of inert tracers to identify discontinuities or mass balance problems;

5. comparing results obtained with different meteorological models;

6. calculating and comparing divergence and/or dimensionless parameters and comparing these with expected ranges;

7. comparing spatial ozone patterns obtained with a grid model vs. observed patterns, and

8. using process analysis to flag contributions made to unexpected ozone concentrations attributable to meteorological factors.

Applying meteorological models over extensive domains with fine scales (i.e., 4-12 km) can be very resource intensive and present data base management problems. We suggest means for reducing such problems.

Air quality inputs are needed for initial conditions and for boundary values at the edges of a modeling domain. There is no satisfactory way to use available air quality observations to specify initial conditions. Thus, States should diminish the importance by beginning a simulation one or more days prior to the period of interest for urban applications and two or more days earlier for regional applications. Nested regional models are the usual preferred means for generating boundary conditions to a portion of the large regional domain which is the focus of an attainment demonstration. If an urban scale model is used, the domain should be large enough so that emissions occurring in the center of the domain just before sunrise remain within the domain until the end of the same calendar day. This should reduce importance of boundary conditions specified for such applications.

Issues related to meteorological models and procedures for accounting for initial and boundary conditions are discussed in Section 14.0.

ES13.0. How Do I Produce Needed Emissions Inputs?

Producing needed emissions inputs requires several steps. First, compile Statewide and then countywide estimates for VOC, NOx and CO emissions for point, area, mobile and biogenic emissions. Second, quality assure the outputs. Third, convert the resulting estimates into speciated, gridded hourly emissions using emissions models. Fourth, once again, quality assure the results. Finally, project gridded, speciated hourly emission estimates to a future year which corresponds to two years prior to the deadline for meeting the NAAQS.

The U.S. EPA has prepared a series of guidelines relating to these steps as a part of the Emission Inventory Improvement Program (EIIP), as well as a guideline for developing

emissions inventories. States should be familiar with these guidelines. States should use the most recent emission estimates commonly available when applying the modeled attainment test used to support an attainment demonstration. If available in a time frame compatible with completing modeling in a timely manner, the National Emissions Trends Inventory (NET) compiled for 1999 is the preferred source of information for State and countywide estimates in portions of the modeling domain for which States who are stakeholders have no better information. Otherwise the NET compiled for 1996, or derived for 1997 or 1998, may be used.

Different means are used to obtain emissions information in the form needed by air quality models. Ideally, location and daily/weekly emission patterns should be directly available for point sources. Spatial distribution of surrogates for activity factors needs to be estimated for major area source categories and for mobile sources in order to estimate gridded emissions. Diurnal and weekly activity patterns are also useful. Current, locally applicable VOC speciation profiles are desirable for point sources and major area source categories. Emission models are then used to characterize emissions from other point sources, stationary area, mobile and biogenic sources. We identify several commonly used emissions models.

Quality assurance of emissions estimates is necessary for an attainment demonstration to be credible. We recommend that it be performed during several stages of the process needed to derive required emission inputs to an air quality model. Comparing inventory estimates made for different studies, computer graphics and comparisons with available, speciated air quality data are useful means for quality assuring emission estimates. Generating emission inputs is addressed more fully in Section 15.0.

ES14.0. How Do I Evaluate Model Performance And Make Use Of Diagnostic Analyses?

In Section 16.0, we describe 7 means for evaluating model performance.

1. Perform a "big picture" assessment using graphics.

2. Use "ozone metrics" in statistical comparisons.

3. Compare predicted and observed precursor or species concentrations.

4. Compare predicted source attribution factors with estimates obtained using observational models.

5. Compare observations and predictions on weekends vs. week days.

6. Compare observed and predicted ratios of indicator species.

7. Use retrospective analyses in which air quality differences predicted with models are

compared with observed trends.

Most of these approaches are only able to address how well a model replicates a past set of observations. While this is useful, the key question is, "how well does a model forecast *changes* in ozone accompanying changes in precursor emissions?". The 5th (weekend/week day comparisons), 6th (use of ratios of indicator species) and 7th (use of retrospective analyses) approaches have the potential to address this key question. However, each requires additional efforts to make certain measurements or to perform additional analyses. We discuss these additional efforts in Sections 5.0 and 16.0.

All of the identified means for evaluating model performance have strengths and weaknesses. Thus, we recommend that as many of these approaches as feasible be used to evaluate model performance. Assessment of whether or not performance is adequate is most properly done by considering evidence produced by all of these evaluation techniques—much the same way as a weight of evidence approach is used in an attainment demonstration.

Diagnostic tests should be applied throughout a modeling analysis. We identify several key stages for use of these tests: (1) during model setup, (2) during model performance evaluation, (3) during the process of choosing/comparing prospective control strategies, (4) to help estimate uncertainty in the resulting air quality projections after selecting a strategy, and (5) to study predictions at specific times and locations in greater depth.

Two types of diagnostic tests are identified: sensitivity tests and process analysis. In designing and evaluating results of diagnostic tests, States should be aware of how models are to be used to support an attainment demonstration. That is, models should be used in a *relative* sense. Thus, diagnostic tests should consider how relative reduction factors (RRF) or other predicted changes in ozone are affected by various factors.

(This page is left blank intentionally)

1.0 Introduction

Readers should note that the U.S. Court of Appeals for the District of Columbia Circuit issued a decision on May 14, 1999 which prohibited enforcement of the 8-hr ozone national ambient air quality standard (NAAQS). Therefore, this draft guidance is being distributed at this time to document the position of the Agency when the Court Decision was issued. Moreover, the concepts developed in the guidance are applicable to other multi-hour standards. When the litigation issues have been addressed and resolved, we will review this draft guidance and revise it as appropriate. Distribution is being made since this document is the product of an extensive dialogue with the stakeholders from the modeling community and since we have had numerous requests for release of the current version. Stakeholders involved in the development of the draft include representatives from State and local governmental agencies, academia, consultants, industry, and environmental organizations in addition to the EPA Regional Offices.

Distribution of this document should not be construed as implementation of the 8-hr standard; it is merely provided for information to interested parties and to document the status of the guidance at the time of the court decision. Thus, we feel it is appropriate to provide this draft guidance document to State and local agencies as well as other requestors. However, no actions should be taken to implement this draft guidance until the litigation issues have been resolved and final guidance has been issued.

1.1 What Is The Purpose Of This Document?

This document has two purposes. The first is to explain how to interpret if results of modeling and other analyses support a conclusion that attainment of the national ambient air quality standard (NAAQS) for 8-hour daily maximum ozone concentrations will occur. The second purpose is to describe how to apply an air quality model to produce results needed to support an attainment demonstration. Part I of this document provides guidance for using results to help demonstrate attainment. Part II provides guidance on how to apply models to produce these results.

With few exceptions, guidance herein should be viewed as recommendations rather than requirements. States may use alternative procedures if these are justified to the satisfaction of the appropriate U.S. EPA Regional Office. Generally, an attainment assessment which leads to greater protection of the environment than that recommended in Part I of this guidance may be used if a State chooses to do so. Although this guidance attempts to address issues that may arise in attainment demonstrations, situations which we have failed to anticipate may occur. These should be resolved on a case by case basis in concert with the appropriate U.S. EPA Regional Office.

1.2 Does The Guidance In This Document Apply To Me?

This guidance applies to all locations which are not attaining the 8-hour NAAQS for

ozone according to data reported to the US EPA's AIRS data base. This includes both "traditional" and "transitional" nonattainment areas. Qualifications for a "transitional" nonattainment area are defined in a 1997 Presidential Directive (Clinton, 1997). Under this directive, States receiving this classification may not have to perform additional modeling under some circumstances. States which have one or more "transitional" nonattainment areas should consult U.S. EPA (1999a). U.S. EPA (1999a) identifies prerequisites to qualify for no additional modeling. That reference also describes what a State needs to do instead, if it elects to do no additional modeling. "Traditional" nonattainment areas and other "transitional" nonattainment areas are subject to the guidance in this document.

State implementation plan (SIP) revisions designed to correct problems meeting the 8-hour NAAQS in traditional nonattainment areas could be due as early as 3 years after an area is designated "nonattainment" (e.g., July 18, 2003, if designation occurs on July 18, 2000).[1] Under this scenario, attainment demonstrations supporting these revisions should be completed by 2002 to allow States sufficient time to complete rulemaking in 2003. Thus, work underlying demonstrations to support a 2003 SIP revision would need to begin no later than 1999.

1.3 How Does The Perceived Nature Of Ozone Affect My Attainment Demonstration?

Guidance for performing attainment demonstrations needs to be consistent with the perceived nature of ozone. In this section, we identify several premises regarding this pollutant. We then describe how the guidance accommodates each.

Premise 1. There is uncertainty accompanying model predictions. "Uncertainty" is the notion that model estimates will not perfectly predict observed air quality at each receptor location, neither now nor in the future. Thus, there will be a distribution of differences between predictions and observations resulting from comparisons on different days at each receptor. Uncertainty arises for a variety of reasons, including limits in the model's formulation which reflect incomplete understanding or a need to make computations tractable, data base limitations and uncertainty in forecasting future determinants of emissions. States should recognize these limitations when preparing their modeled attainment demonstrations, as should those reviewing the demonstrations.

We recommend several qualitative means for recognizing model limitations and resulting uncertainties when preparing an attainment demonstration. First, we recommend using models in a relative sense in concert with observed air quality data (i.e., taking the ratio of future to present predicted air quality and multiplying it times a monitored design value). As described later, we believe this approach should reduce some of the uncertainty attendant with using absolute model predictions alone. Second, we recommend that a modeling analysis be preceded by analyses of

---

[1]Timing may differ for "traditional" and "transitional" nonattainment areas. See U.S. EPA (1999a) for further discussion of "transitional" nonattainment areas.

available air quality, meteorological and emissions data to gain a qualitative description of an area's nonattainment problem. Such a description should be used to help guide a model application and may provide a reality check on the model's predictions. Third, we offer the option for States to use several model outputs, as well as other analyses besides models to provide corroborative evidence concerning adequacy of a proposed strategy for meeting the NAAQS. Outcomes of modeling and other analyses are weighed to determine whether or not the resulting evidence suggests a proposed control strategy is adequate to meet the NAAQS. Finally, we identify several activities/analyses which States could undertake, if they so choose, to better apply models and corroborative approaches in subsequent reviews/analyses of a control strategy. These subsequent reviews may be useful for determining whether a SIP is working as expected. A State has the responsibility to prepare a subsequent SIP revision, if the U.S. EPA finds that the SIP is substantially inadequate to achieve the NAAQS.

Premise 2. Resource intensive approaches may often be needed to support an adequate attainment demonstration. This follows from the regional nature of ozone concentrations approaching 0.08 ppm in large portions of the U.S. While we believe that regional reductions in NOx emissions should reduce ozone in much of the eastern U.S., concentrations approaching the .08 ppm level specified in the NAAQS will affect local strategies needed to attain the NAAQS in the remaining nonattainment areas.

If regionality is a problem, this guidance recommends using regional modeling domains. Regional modeling applications require coordination, quality assurance and management of data bases covering large areas of the country. Resources used to run recommended models for generating meteorological and emissions inputs and the air quality model itself can be substantial. States facing the need to develop an attainment demonstration requiring resource intensive techniques may wish to consider pooling resources in some manner. Examples might include delegating responsibilities for certain parts of the analyses to a single State which can "specialize" in that kind of analysis. Another example might be formation of a regional center of some kind to perform analyses as directed by its client group of States.

Premise 3. There will be a widespread need to use nested regional models. Available air quality data suggest ozone concentrations approach levels specified in the NAAQS throughout much of the eastern U.S. and in large parts of California. Near nonattainment areas, more detailed attention may need to be paid to atmospheric mixing of nearby emissions than is necessary for emissions in locations which are more remote from the nonattainment areas. This is consistent with use of nested regional models.

This guidance identifies several *modeling systems*[2] with nesting capabilities, including

---

[2]A modeling system includes a chemical model, an emissions model and a meteorological model. Terms, such as this one, which are introduced using italics are defined more fully in a glossary at the back of this guidance. "Modeling system" and "air quality model" are used interchangeably. "Air quality model" means "modeling system" in this guidance.

one for which the US EPA will provide user support (MODELS3/CMAQ) (U.S. EPA, 1998a). Support includes documentation, ready access to the code, training, updates and limited troubleshooting. We believe it is premature to identify MODELS3/CMAQ or any other nested regional modeling system as the "guideline model" for ozone. States may use MODELS3/CMAQ or an alternate modeling system provided certain criteria, identified in this guidance, are met. These criteria apply equally to MODELS3/CMAQ and alternative air quality model(s). The guidance also provides recommendations for developing meteorological, air quality and emissions inputs used in nested regional modeling systems, and makes suggestions for quality assuring inputs and evaluating performance of emissions, meteorological and air quality models.

Premise 4. Problems posed by high ozone, $PM_{2.5}$[3] and regional haze share several commonalities. Ozone formation and formation of secondary particulates result from several common reactions and reactants. Secondary particulates are a major part of $PM_{2.5}$. Often similar sources contribute precursors to both ozone and $PM_{2.5}$. In some regions of the U.S., high regional ozone and secondary particulates have been observed under common sets of meteorological conditions. Reducing $PM_{2.5}$ is the principal controllable means for improving regional visibility. U.S. EPA policy is to encourage "integration" of programs to reduce ozone, $PM_{2.5}$ and regional haze to ensure they do not work at cross purposes and to foster maximum total benefit for lower costs.

Integration of strategies to reduce ozone, $PM_{2.5}$ and regional haze is complicated by different dates likely needed for SIP revisions (e.g., 2003 for ozone, circa 2007-2008 for $PM_{2.5}$ and regional haze, etc.). One reason for a subsequent review of a strategy selected to meet the ozone NAAQS is to check its compatibility with plans (whose details become known later) to meet goals for $PM_{2.5}$ and regional haze. This guidance identifies activities which could yield useful information for a subsequent review to help States prepare for such a check if they so choose. Use of air quality models, such as MODELS3/CMAQ, which also have the capability of considering effects of control strategies on $PM_{2.5}$ and regional haze is desirable.

1.4 What Topics Are Covered In This Guidance?

This guidance addresses two broad topics: Part I, "How do I use results of models and other analyses to help demonstrate attainment?", and Part II, "How should I apply air quality models to produce results needed to help demonstrate attainment?". Part I is divided into 6 sections (i.e., Sections 2.0-7.0). Part II consists of 10 sections (Sections 8.0-17.0).

Section 2.0 contains an overview of the procedure we recommend for using results to help demonstrate attainment of the 8-hour ozone NAAQS. The recommended approach is to first use an air quality model to estimate current and future ozone concentrations. Next, use the predicted relative changes in ozone in concert with measured data to estimate future ozone

---

[3] $PM_{2.5}$ are particles having aerodynamic diameters less than or equal to 2.5 micrometers.

concentrations. We refer to this exercise as a "*modeled attainment test*". If the test is passed and a similar *screening test*, applied at selected locations without monitors, is also passed, these outcomes suggest attainment will occur if the simulated control strategy is adopted. States have an option to use a suite of model predictions as well as several additional data analyses for corroborating conclusions reached with the modeled attainment and screening tests. Corroborative analyses use air quality and emissions data plus additional interpretation of model results. Results of the modeled attainment/screening tests and corroboratory analyses are considered together in a *weight of evidence determination* to assess whether or not a proposed control strategy is likely to be successful in meeting the NAAQS. "Weight of evidence" may be used either to require more or to require fewer control measures than the modeled attainment test suggests is necessary.

Section 3.0 describes the recommended modeled attainment test and screening test in detail. The Section includes examples illustrating use of the modeled attainment and screening tests we recommend.

Section 4.0 describes how a weight of evidence determination should be performed, if a State chooses to use evidence produced by corroboratory analyses to complement results of a modeled attainment test and screening test. A weight of evidence determination consists of modeling plus a series of other core corroborative analyses plus additional, optional analyses. Modeling is generally the most reliable means for estimating future ozone concentrations, because it integrates a diverse set of information into a coherent description of ozone formation, transport and decay. Evidence that a model is able to reproduce a detailed observed data base makes its predictions more compelling. Modeling should always be included as part of a weight of evidence determination. Each core analysis in a weight of evidence determination is identified, along with conditions which lend credibility to its outcome. Outcomes which are consistent with concluding that a proposed control strategy will work are also identified. Several examples of optional analyses are also provided, along with recommendations for accompanying documentation.

Section 5.0 identifies several data gathering activities and analyses which States could undertake to enhance modeling and corroborative analyses to support subsequent reviews. None of these activities is required for the current ozone SIP revision. However, they would increase credibility of the modeling/analysis exercise. A subsequent review will be desirable to diagnose why a strategy is or isn't working, or to relate the chosen strategy to others which are later considered to reduce $PM_{2.5}$ or regional haze. The less extensive a data base underlying a *modeled attainment demonstration*, the greater the potential need for a subsequent review.

Section 6.0 identifies the necessary documentation describing the analyses used to demonstrate attainment of the ozone NAAQS.

Section 7.0 lists the references cited in Part I and in this introduction (Section 1.0).

Part II ("How should I apply air quality models to produce results needed to help demonstrate attainment?") begins in Section 8.0 with an overview of the topics to be covered.

Section 9.0 identifies a series of meteorological, emissions and air quality data analyses which should be undertaken to develop a qualitative description of an area's nonattainment problem prior to a model application. As we describe, this qualitative description should be used to guide the subsequent model application.

Section 10.0 identifies the need for a modeling/analysis protocol. We also discuss the protocol's function as well as what subjects should be addressed in the protocol.

Section 11.0 addresses what should be considered in choosing a model to support the attainment demonstration of the ozone NAAQS. Several criteria are identified for accepting use of a model for this purpose.

Section 12.0 provides guidance for selecting suitable episodes to model for an ozone attainment demonstration. Topics include a discussion of the form of the NAAQS and its resulting implications for episode selection.

Section 13.0 identifies factors which should be considered in choosing a model domain and horizontal and vertical resolution for the model application.

Section 14.0 addresses how to develop meteorological inputs as well as initial and boundary air quality data for use in a modeling exercise supporting an attainment demonstration. Topics covered include use of dynamic meteorological models, four dimensional data assimilation, relationship to domain size and the need for "ramp-up" days.

Section 15.0 discusses how to develop appropriate emissions estimates for use in the selected air quality model. Topics include use of available inventory estimates, quality assurance, application of emissions models and estimating future emissions.

Section 16.0 treats the topics of model performance evaluation and use of diagnostic analyses.

The guidance concludes with Section 17.0, which lists references cited in Part II, and a glossary of important terms which may be new to some readers.

Part I. How Do I Use Results Of Models And Other Analyses To Help Demonstrate Attainment?

(This page is left blank intentionally)

2.0 What Is A Modeled Attainment Demonstration?--An Overview

A *modeled attainment demonstration* consists of (a) analyses which estimate whether selected emissions reductions will result in ambient concentrations that meet the NAAQS, and (b) an identified set of measures which will result in the required emissions reductions. As noted in Section 1.0, this guidance focuses on the first component of an attainment demonstration--interpretation and conduct of analyses to estimate the amount of emission reduction needed to reduce ozone concentrations to a level which is consistent with meeting the NAAQS. Emission reduction strategies should be simulated by reducing emissions from specific source categories rather than through broad "across-the-board" reductions from all sources.

States should estimate the amount of emission reduction needed to demonstrate attainment by using a *modeled attainment test* plus using a *screening test* at selected locations without an ozone monitor. In addition to these tests, a State may consider a broader set of model results plus perform a set of other corroboratory analyses to determine whether the "weight of evidence" produced by the tests and additional analyses indicates that a proposed emission reduction will lead to attainment of the NAAQS.

2.1 What Is The Recommended Modeled Attainment Test?--An Overview

A *modeled attainment test* is an exercise in which an air quality model is used to simulate current and future air quality. If future estimates of an ozone design value are ≤ 84 ppb, the test is passed. Our recommended test is one in which model estimates are used in a "relative" rather than "absolute" sense. That is, we take the ratio of the model's future to current predictions at ozone monitors. We call each of these site-specific ratios, *relative reduction factors*. Future ozone design values are estimated at existing monitoring sites by multiplying a modeled relative reduction factor at locations "near" each monitor times the observed monitor-specific ozone design value. The resulting predicted site-specific "future design value" is compared to 84 ppb. If all such future site-specific design values are ≤ 84 ppb, the test is passed.

The modeled attainment test we recommend predicts whether or not all observed future design values will be less than or equal to the concentration level specified in the NAAQS for ozone under meteorological conditions similar to those which have been simulated. By itself, the test makes no statement about future ozone at locations where there is no nearby monitor. Thus, we require a supplementary screening analysis to identify other locations where passing the test might be problematic if monitoring data were available. Like the test itself, this supplementary *screening test* is described more fully in Section 3.0. Briefly however, it entails the following:

-identification of areas in the modeling domain where "absolute" predicted 8-hour daily maxima are consistently greater than those predicted in the vicinity of any monitor site,

and

-computation of relative reduction factors for each identified unmonitored area with high predicted ozone. These factors are then multiplied by the *areawide design value* to obtain an estimated future design value for each such location.

2.2 What Does A Recommended Weight Of Evidence Determination Consist Of?--An Overview

As we note later in Section 9.0, States should always perform complementary analyses of air quality, emissions and meteorological data, and consider modeling outputs other than the results of the attainment and screening tests. Such analyses are instrumental in guiding the conduct of an air quality modeling application. Sometimes, results of corroboratory analyses may be used in a *weight of evidence determination* to conclude that attainment is likely despite modeled results which do not quite pass the attainment and/or screening tests. The further attainment or screening tests are from just being passed, the more compelling contrary evidence produced by corroboratory analyses must be to draw a conclusion differing from that implied by the test results. If a conclusion differs from the outcome of the modeled tests, the need for subsequent review (several years hence) with more complete data bases is increased. If either test is failed by a wide margin, we doubt that the more qualitative arguments made in a weight of evidence determination can be sufficiently convincing to conclude that the NAAQS will be attained. Table 2.1 contains rules of thumb for assessing whether a weight of evidence determination may be appropriate.

Table 2.1. Guidelines For Using A Weight of Evidence Determination

Results of Modeled Attainment Test*	May A Weight Of Evidence Determination Be Used?
Future Design Value ≤ 84 ppb, all sites	Yes
Future Design Value 85 - 89 ppb at one or more sites	Yes
Future Design Value ≥ 90 ppb, at one or more sites	Not ordinarily. More qualitative results are unlikely to reach a conclusion differing from the outcome of an attainment or screening test.

* Includes calculations at screening sites, if applicable

In a weight of evidence (WOE) determination, States should review results from several diverse types of analyses, including results from the modeled attainment test and, if applicable, the screening test. States should next note whether or not results from each of these analyses support a conclusion that the proposed strategy will meet the air quality goal. States should then weigh each type of analysis according to its credibility as well as its ability to address the question being posed (i.e., is the strategy adequate for meeting the ozone NAAQS by a defined deadline?). Next, conclusions derived in the two preceding steps are combined to make an overall assessment of whether meeting the air quality goal is likely. This last step is a qualitative one involving some subjectivity. If it is concluded that a strategy is inadequate to demonstrate attainment, a new strategy is selected for review, and the process is repeated. States should provide a written rationale documenting how and why the conclusion is reached regarding adequacy of the final selected strategy.

Results obtained with air quality models are an essential part of a weight of evidence determination and should ordinarily be very influential in deciding whether the NAAQS will be met. This follows from including *ability to address the question being posed* as one of two criteria for weighing results from different analyses and from a model's ability to integrate information from scientific theory and observed data. If the modeled attainment and screening tests are passed, this supports a hypothesis that the strategy is adequate. This information is included as one of several elements in a weight of evidence determination to assess the strategy's adequacy. The further model results are from passing the modeled attainment or screening test, the more compelling results from other analyses have to be for a control strategy to demonstrate attainment. If either the modeled attainment test or screening test produces one or more estimated site-specific future design values $\geq$ 90 ppb, it is doubtful that other evidence will be sufficiently convincing to conclude that the NAAQS will be attained. States should ordinarily consider a revised control strategy.

2.3 Why Should A Model Be Used In A "Relative" Sense And Why May Corroboratory Analyses Be Used In A Weight Of Evidence Determination?

The procedure we recommend for estimating needed emission reductions differs from that in past guidance for ozone in two major respects (U.S.EPA, 1996). First, we recommend a modeled attainment test in which model predictions are used in a relative rather than absolute sense. Second, the role of the weight of evidence determination, when used, has been expanded. That is, results can now be used as a basis for requiring emission reductions greater than those implied by the modeled attainment test as well as a rationale for concluding that a control strategy will meet the NAAQS, even though the modeled attainment or screening test is not quite passed. There are several reasons why we believe these changes are appropriate.

1. The form of the 8-hour NAAQS necessitates such an attainment test. The 8-hour NAAQS for ozone requires the 4th highest 8-hour daily maximum ozone concentration, averaged over 3

consecutive years, to be ≤ 0.08 ppm at each monitoring site[4]. The feature of the NAAQS requiring averaging over 3 years presents difficulties using the resource-intensive episodic models we believe are necessary to capture spatially differing, complex non-linearities between ambient ozone and precursor emissions. That is, it is difficult to tell whether or not a modeled exceedance obtained on one or more days selected from a limited sample of days is consistent with meeting the NAAQS. To do so would require modeling many days and, perhaps, many strategies. This problem is reduced by using the *monitored* design value, calculated consistently with the form of the NAAQS, as an inherent part of the modeled attainment test.

2. Current design values for the 8-hour NAAQS are generally closer to the concentration specified in the NAAQS than is true for the 1-hour NAAQS for ozone. A review of 1994-96 ozone data reported in the U.S. EPA's AIRS data base suggests that most nonattainment areas have design values that are within about 40 ppb of the concentration specified in the 8-hour NAAQS. Thus, the "signal to noise ratio" in model applications related to the 8-hour NAAQS is lower than is the case for the 1-hour NAAQS, if we continue to use absolute model predictions as the basis for the modeled attainment test. This follows since the ozone concentration we are trying to reduce to the level of the NAAQS is closer to that level than is true for the 1-hour NAAQS. Therefore, difficulties posed by model uncertainty may be greater for the 8-hour NAAQS than for the 1-hour NAAQS for ozone.

3. Starting with an observed rather than modeled concentration as the base value subject to improvement reduces problems in interpreting model results. This follows for two reasons. First, if a model under (or over) predicts an observed daily maximum concentration, the appropriate target prediction is not so clear as might be desired. For example, if an 8-hour daily maximum ozone concentration of 120 ppb were observed and a model predicted 100 ppb on that day, should the target for the day, nevertheless, be 84 ppb? Although good model performance remains a prerequisite for use of a model in an attainment demonstration, problems posed by disagreements on an individual day are reduced by the new procedure. Second, as described later, we have found that relative reduction factors reflecting predicted 8-hour daily maxima averaged over several days are insensitive to the magnitude of a predicted current 8-hour daily maximum concentration averaged over several days, unless the prediction is below about 70 ppb. This finding may facilitate using days with intensive data bases (for model evaluation) even though such days are not among the ones with the very highest observed concentrations of ozone.

4. Model results and projections will continue to have associated uncertainty. The procedure we recommend recognizes this by allowing use of modeling plus other analyses to determine whether weight of available evidence supports a conclusion that a proposed emission reduction will suffice to meet the NAAQS.

---

[4]See 40CFR Part 50.10, Appendix I, paragraph 2.3. Because of the stipulations for rounding significant figures, this equates to a modeling target of ≤ 84 ppb. Because non-significant figures are truncated, a modeling estimate < 85 ppb is equivalent to ≤ 84 ppb.

5. Focusing the modeled attainment test on monitoring sites could result in control targets which are too low if the monitoring network is limited or poorly designed. We recommend using a test for selected locations without monitors. This exercise provides a screening test for identifying a possible need for more controls despite passing the modeled attainment test. As noted in Table 2.1, a weight of evidence determination may also be used. A weight of evidence determination includes several modeling results which are more difficult to relate to the form of the NAAQS. These results address relative changes in the frequency and intensity of high modeled 1-hour concentrations or high 8-hour daily maxima on the sample of days selected for modeling. If corroboratory analyses produce strong evidence that a control strategy is unlikely to meet the NAAQS, a weight of evidence determination may be used to conclude the strategy is inadequate, even if the modeled attainment test is passed.

Recommendations. States should estimate emission reductions needed to demonstrate attainment using a modeled attainment test and a screening test applied at selected sites without monitors. The modeled attainment test should use model predictions in a relative sense to compute relative reduction factors associated with a strategy. These factors should be multiplied by monitored design values at different monitors to estimate future design values to compare with 84 ppb. When necessary, the modeled attainment test should be supplemented with a screening test which applies similar procedures using the areawide design value at selected locations without monitors.

States should undertake complementary analyses of air quality, meteorological and emissions data. These additional analyses are needed to design and focus modeling which underlies the attainment test. Provided results of the attainment and screening tests are not failed by a wide margin, States may use evidence produced by corroborative analyses together with results of the tests in a weight of evidence determination. A weight of evidence determination includes the modeled attainment and screening test results, plus results of additional model outputs plus other analyses of air quality, meteorological and emissions data. A weight of evidence analysis may be used either to increase or decrease emission reductions identified as sufficient to meet the NAAQS by a modeled attainment test.

(This page is left blank intentionally)

3.0 What Is The Recommended Modeled Attainment Test?

In Section 2.0, we provided an overview of the recommended modeled attainment test. However, there are several decisions which must be made before the recommended test can be applied. In this Section, we identify a series of issues regarding selection of inputs to the test, and recommend solutions. We next describe how to apply the test and illustrate this with examples. We then identify some implications resulting from the test. We conclude with a further discussion of a screening test recommended for locations without monitors for which predictions are consistently higher than any near a monitoring site.

Equation (3.1) describes the recommended modeled attainment test, applied near monitoring site I.

$$(DVF)_I = (RRF)_I (DVC)_I \qquad (3.1)$$

where

$(DVC)_I$ = the current design value (e.g., 1997-99) monitored at site I, ppb

$(RRF)_I$ = the relative reduction factor, calculated near site I, unitless.
The relative reduction factor is the ratio of the future 8-hour daily maximum concentration predicted near a monitor (averaged over several days) to the current 8-hour daily maximum concentration predicted near the monitor (averaged over the same several days), and

$(DVF)_I$ = the estimated future design value for the time attainment is required, ppb.

Equation (3.1) looks simple enough. However, several issues must be resolved before applying it.

(1) How is a "site-specific" current design value ($(DVC)_I$) calculated?

(2) For which of the 3 years to be used in the future to assess attainment from monitored data, should future emissions be calculated for the modeled attainment test?

(3) In calculating the $(RRF)_I$, what do we mean by "near" site I?

(4) Several surface grid cells may be "near" the monitor, which one(s) of these should be used to calculate the $(RRF)_I$?

(5) Should any days be excluded when computing a relative reduction factor?

The preceding questions can be lumped under a single question, "how do I select appropriate

inputs for the modeled attainment test?"

3.1 How Do I Select Appropriate Inputs For The Modeled Attainment Test?

Calculating the current site-specific design value $(DVC)_I$. The modeled attainment test is linked to the form of the 8-hour NAAQS for ozone through use of current monitored design values, calculated consistently with the form of the NAAQS. The current site-specific monitored design value $(DVC)_I$ is calculated at each monitor by noting the 4th highest daily maximum concentration in each of 3 consecutive years. The arithmetic mean of these 3 values is then computed. Table 3.1 illustrates the procedure (i.e., note that fractions of a ppb get truncated).

Table 3.1. Example Monitoring Data for Nonattainment Area "A"

(1) Monitoring Site Number I	(2) 4th High 8-hr Daily Max., Year 1	(3) 4th High 8-hr Daily Max., Year 2	(4) 4th High 8-hr Daily Max., Year 3	(5) Average 4th High 8-hr Daily Max.
Monitor 1	94 ppb	87 ppb	91 ppb	90.67 = 90 ppb
Monitor 2	89 ppb	81 ppb	82 ppb	84 ppb
Monitor 3	94 ppb	83 ppb	81 ppb	86 ppb

The site-specific current design values ($(DVC)_I$) used in the recommended modeled attainment test are shown in column (5) of the table. The areawide design value (used in the screening test described in Section 3.4) corresponds to the highest of the average 4th high daily maximum ozone concentrations. Thus, the areawide design value for Nonattainment Area "A" is 90 ppb.

What if a monitor has less than 3 years of complete data?

We believe it is important to consider as much available air quality data as defensible in the modeled attainment test. Thus, States should include monitoring sites having less than 3 years of complete data for the current (e.g., 1997-99) period. As described below, *for use in the modeled attainment test*, a State should calculate current design values at sites having observations during two or more years during the current period. The following order of preference should be used for calculating the current design value for each site in the modeled attainment test:

1. Three years of complete (i.e., $\geq$ 74.5% days during the ozone season have valid 8-hour daily maxima) data are available in the current period. Calculate the current design value as shown in Table 3.1;

2. Three consecutive years of complete data are available, but only two of these are within the current period (e.g., complete data are available for 1996-98, but not for 1997-99). However, year 3 (e.g., 1999) of the current period has some observations at the site in question. Calculate the current design value as shown in Table 3.1 for

(a) 1996-1998 (in this example), and

(b) 1997-1999 (in this example), using the 4th high 8-hour daily maximum ozone concentration observed in year 3 (1999, in this example).

Use the higher of "(a)" or "(b)" as the current design value.

3. If the conditions in neither "1" nor "2" are met at a site, but the site has some observations in 2 or 3 of the years in the current period, choose the higher of "(c)" or "(d)". If neither "(c)" nor "(d)" exceeds 84 ppb, do not calculate a current design value for this site nor include it in the modeled attainment test. Under these circumstances, including the site in the test could result in an overly optimistic estimate of a future design value.

(c) Calculate the current design value by taking the mean of the 4th highest 8-hour daily maximum over 2 years at a site with 2 years' observations.

(d) Calculate the current design value by taking the mean of the 4th highest 8-hour daily maximum over 3 years at a site with 3 years' observations.

Choosing the "current period" to use in the attainment test.

Over consecutive overlapping 3-year periods, variability in observed design values is heavily influenced by meteorological variations. A State should seek some assurance that "current design values" used in the attainment test do not reflect a period in which conditions for high ozone were unfavorable at the monitoring site. Accordingly, States should review monitored data from (a) the 3-year period "straddling" the year represented by the most recent available emissions inventory (e.g., 1995-97, for a 1996 inventory), and (b) the 3-year period used to designate an area "nonattainment"[5]. The current design value used in the modeled attainment and screening tests is the higher of the two estimates obtained from (a) and (b). For the modeled attainment test, this choice should be made on a monitor by monitor basis. Selection of appropriate current design values is illustrated below.

Example 3.1

Given: (1) An area is designated "nonattainment" for the 8-hour ozone NAAQS on the basis of

---

[5]For later, subsequent reviews, (b) becomes instead the most recent available 3-year contiguous period of ozone observations

1997-1999 monitored data.
(2) The most recent inventory reflects 1996 emissions.
(3) For purposes of illustration, suppose the area contains only 4 ozone monitors.

Find: The appropriate site-specific current design values to use in the modeled attainment test.

Solution: Since the most recent inventory reflects 1996, we need to examine monitored design values for 1995-1997 and compare these to design values for 1997-1999. Choose the highest of the two design values at each site. These are the values for DVC in the modeled attainment test. The procedure is shown in Table 3.2.

Table 3.2. Example Illustrating Selection Of Current Design Values

Monitor	1995-97 Design Value, ppb	1997-99 Design Value, ppb	Current Design Value (DVC) Used In The Modeled Attainment Test, ppb
1	72	80	80
2	94	91	94*
3	90	92	92
4	86	81	86

* Areawide design value used in the screening test (see Section 3.4)

Choosing a year to project future emissions. States should project future emissions to the year in which all control measures needed to meet the NAAQS are in place (U.S. EPA(1999a)). Since the NAAQS incorporates observations from three years, the projection year should be the first year of the 3 year period used to judge attainment. For example, if the year by which attainment is required were 2007, air quality monitoring data from 2005-2007 would ultimately be used to judge whether the NAAQS has been attained. Thus, in this example, the projection year used in the modeled attainment test should be 2005. As noted in Section 6.0, accompanying documentation should show that emissions remain at or below amounts projected to the initial year of the 3-year period.

Identifying surface grid cells near a monitoring site. There are three reasons why we believe it is appropriate, in the modeled attainment test, to consider cells "near" a monitor rather than just the cell containing the monitor. First, one consequence of a control strategy may be migration of a predicted peak. If a State were to confine its attention only to the cell containing a monitor, it might underestimate the RRF (i.e., overestimate the effects of a control strategy).

Second, we believe that uncertainty in the model's formulation and inputs is consistent with recognizing some leeway in the precision of the predicted location of daily maximum ozone concentrations. Finally, standard practice in defining a gridded modeling domain is to start in the Southwest corner of the domain, and reckon grid cell location from there. This is often, and indeed should be, many kilometers from monitoring sites in a modeled traditional nonattainment area. Considering several cells "near" a monitor rather than the single cell containing the monitor diminishes the likelihood of quirks in the test's results resulting from geometry of the superimposed grid system.

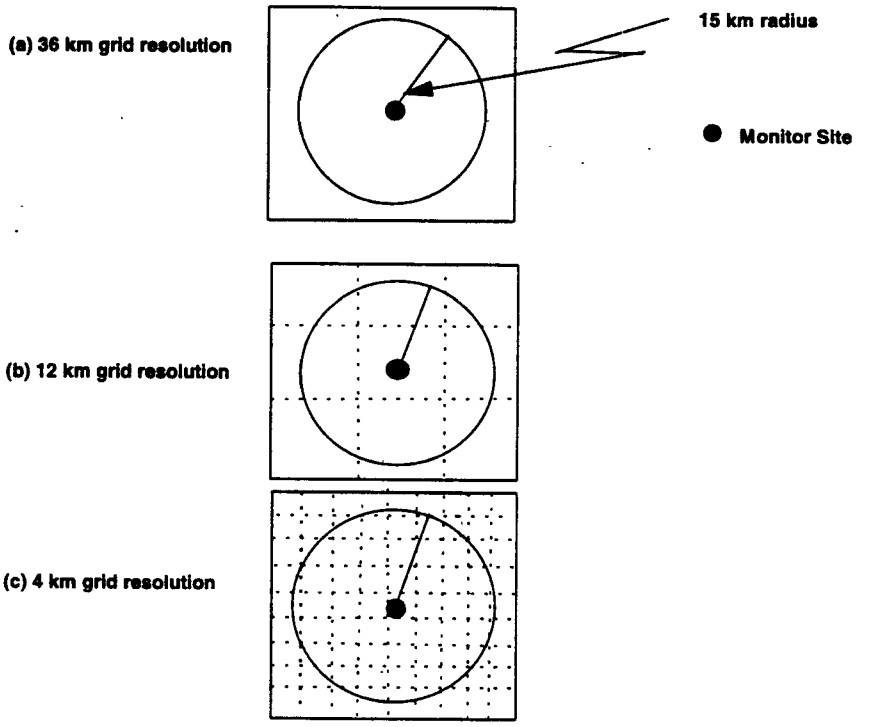
Earlier guidance (U.S. EPA(1996)) has identified 15 km as being "near" a site. This is also consistent with the broad range of intended representativeness for urban scale ozone monitors identified in 40CFR Part 58, Appendix D.

For ease in computation, States may assume that a monitor is at the center of the cell in which it is located and that this cell is at the center of an array of "nearby" cells. As shown in Figure 3.1, the number of cells considered "nearby" (i.e., within about 15 km of) a monitor is a function of the size of the grid cells used in the modeling. Table 3.3 provides a set of default recommendations for defining "nearby" cells for grid systems having cells of various sizes. Thus, if one were using a grid with 4 km grid cells, "nearby" is defined by a 7 x 7 array of cells, with the monitor located in the center cell. States may consider presence of topographic features, demonstrated mesoscale flow patterns (e.g., land/sea, land/lake interfaces) or other factors to refine our default definitions for the array of "nearby" grid cells, provided the justification for doing so is documented.

Table 3.3. Default Recommendations For Nearby Grid Cells Used To Calculate RRF's

Size of Individual Cell, km	Size of the Array of Nearby Cells, unitless
≤5	7 x 7
>5 - 8	5 x 5
>8 - 15	3 x 3
>15	1 x 1

Figure 3.1. Relationship Between Grid Resolution and Grid Cells Considered to be in the Vicinity of a Monitor



Choosing model predictions to calculate a relative reduction factor ($(RRF)_I$) near a monitor. Two decisions need to be made. First, given that a model application produces a time series of estimated 1-hour ozone concentrations (which can be used to calculate running 8-hour averages), what values should be chosen from within the time series? We recommend choosing predicted 8-hour daily maximum concentrations from each modeled day (excluding "ramp-up" days) for consideration in the modeled attainment test. The 8-hour daily maxima should be used, because they are closest to the form of concentration specified in the NAAQS.

The second decision that needs to be made is, "which one(s) of the 8-hour daily maxima predicted in cells near a monitor should we use to calculate the RRF?" We recommend choosing the nearby grid cell with the highest predicted 8-hour daily maximum concentration with current emissions for each day considered in the test, and the grid cell with the highest predicted 8-hour daily maximum concentration with the future emissions for each day in the test. Note that, on any given day, the grid cell chosen with the future emissions need not be the same as the one chosen with current emissions.

We believe selecting the maximum (i.e., peak) 8-hour daily maxima on each day for subsequently calculating the relative reduction factor (RRF) is preferable for several reasons. First, it is likely to reflect any phenomenon which causes peak concentrations within a plume to migrate as a result of implementing controls. Second, it is likely to take better advantage of data produced by a finely resolved modeling analysis.

The relative reduction factor (RRF) used in the modeled attainment test is computed by taking the ratio of the mean of the selected daily maximum predictions in the future to the mean of the selected 8-hour daily maximum predictions with current emissions. The procedure is illustrated in Example 3.2.

Example 3.2

Given: (1) Four primary days have been simulated using current and future emissions;

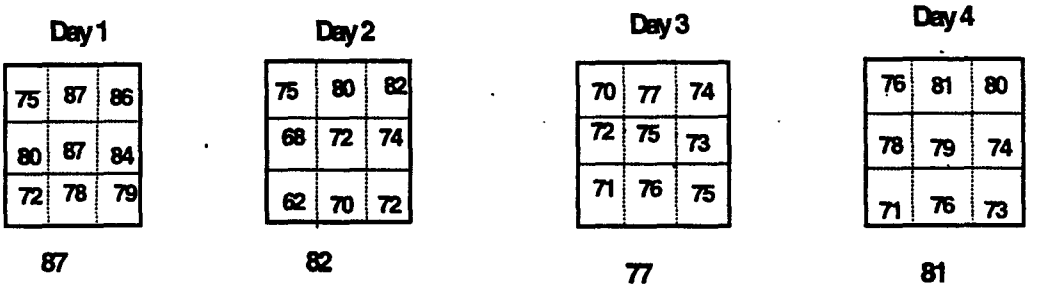
(2) The horizontal dimensions for each surface grid cell are 12 km x 12 km;

(3) Figure 3.2 shows predicted 8-hour daily maximum ozone concentrations in each of the 9 cells "near" a monitor with (a) future emissions, and (b) current emissions.

Find: The site-specific relative reduction factor for monitoring site I, $(RRF)_I$

Figure 3.2. Choosing Predictions to Estimate RRF's

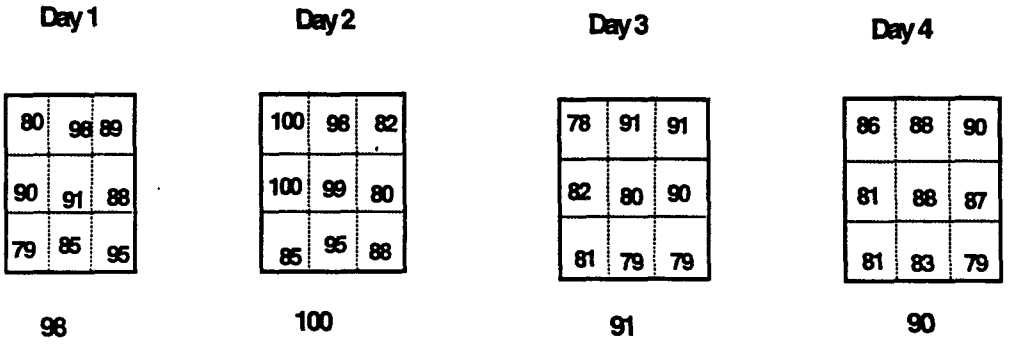
(a) Predictions With Future Emissions



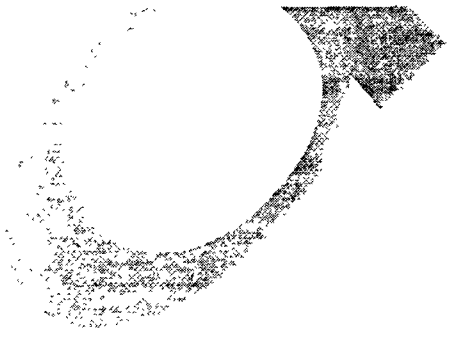
Future Mean Peak 8-hr Daily Max. = (87 + 82 + 77 + 81) / 4 = 81 ppb

---

(b) Predictions With Current Emissions



Current Mean Peak 8-hr Daily Max. = (98 + 100 + 91 + 90) / 4 = 94 ppb



Solution:

(1) For each day and for both current and future emissions, identify the peak 8-hour daily maximum concentration predicted near the monitor. Since the grid cells are 12 km, a 3 x 3 array of cells is considered "nearby" (see Table 3.3). The numbers appearing beneath each 3 x 3 array in Figure 3.2 are the peak nearby concentrations for each day.

(2) Compute the mean peak 8-hour daily maximum concentration for (a) future and (b) current emissions.

Using the information in Figure 3.2,

(a) (Mean peak 8-hr daily max.)$_{future}$ = (87 + 82 + 77 + 81)/4 = 81.75 = 81 ppb
(Note that we have truncated the insignificant figures), and

(b) (Mean peak 8-hr daily max.)$_{current}$ = (98 + 100 + 91 + 90)/4 = 94 ppb

(3) The relative reduction factor near site I is

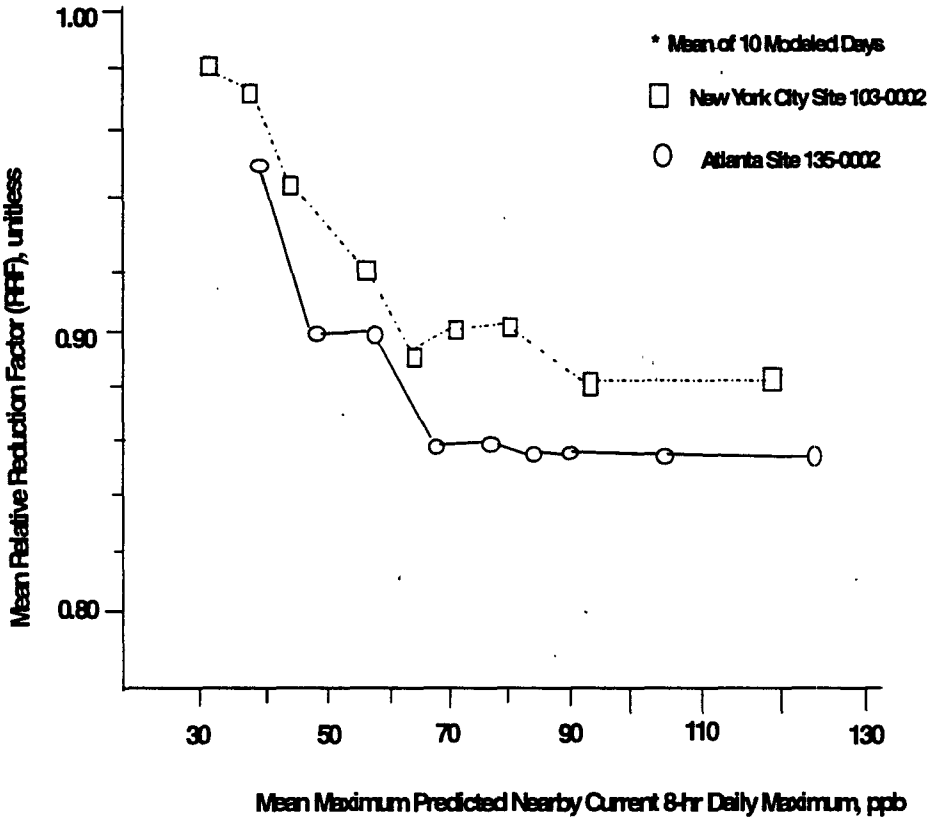
$(RRF)_I$ = (mean peak 8-hr daily max.)$_{future}$/(mean peak 8-hr daily max.)$_{current}$

= 81/94 = 0.86

Limiting modeled 8-hour daily maxima chosen to calculate RRF. On any given modeled day, meteorological conditions may not be similar to those leading to high concentrations (i.e., near the site-specific design value) at a particular monitor. If ozone predicted near a monitor on a particular day is very much less than the design value, the model predictions for that day could be unresponsive to controls (e.g., the location could be upwind from most of the emissions in the nonattainment area on that day). Using equation (3.1) could then lead to an erroneously high projection of the future design value.

Figure 3.3 shows results from a study in which we modeled current and future emissions for 90 days during 1995 using a grid with 12 km x 12 km cells and 7 vertical layers. One purpose of the study was to assess the extent to which a relative reduction factor (RRF) is dependent on the magnitude of predicted current 8-hour daily maxima. We examined RRF's computed near each of 158 monitoring sites in the eastern half of the United States. Sites reflect a variety of surroundings and reductions in surrrounding VOC and NOx emissions to varying degrees. The curves depicting the relationship between mean current 8-hour daily maximum concentrations and RRF averaged over 10 days for the two sites shown in the Figure are typical. Generally, the RRF is not a strong function of the predicted current 8-hour daily maximum ozone concentration *averaged over several days* when these averages are $\geq$ 70 ppb. We would expect relationships like those in Figure 3.3 to be more variable if they reflected averages over only 1-2

Figure 3.3. Mean Relative Reduction as a Function of Mean Predicted Current 8-hour Daily Maxima*



days. Thus, it is better to simulate several days so that RRF values are less likely to be affected by how closely a model's predictions match observed 8-hour daily maxima at individual sites on a given day.

The episode selection procedure recommended in Section 12.0 should help focus modeling on days with observed concentrations near a nonattainment area's design value. Nevertheless, there will inevitably be some modeled days where the predicted 8-hour daily maximum ozone concentrations near a monitoring site do not reflect conditions leading to observations near its design value. To illustrate with a simple example, consider a city with 2 monitors, one north of the city and one south. We would expect the situation of the city to observe high ozone, at or near the design value on the selected days with wind blowing to the north. However, on days when the wind is to the south, the northern site may see little benefit from the control strategy. If local emissions are influential in affecting observed concentrations, we would expect to predict concentrations well below the northern site's design value on a day with winds to the south. Presumably, there would be several such modeled days, since the analysis needs to provide assurance that a strategy will suffice to meet the NAAQS at all sites in the nonattainment area, including the site south of the city.

As a rule of thumb to avoid overestimates of future design values, we recommend excluding some modeled days from consideration when the modeled attainment test is applied near a monitoring site. More specifically, states need not consider any day for which the predicted current maximum 8-hour daily maximum concentration at a nearby grid cell is < 70 ppb. Example 3.3 illustrates what to do if low current predictions occur near a monitor on a day (e.g., as might happen if the monitor is "upwind" on that day).

Example 3.3

Given: The same simulations as performed in Example 3.2 yield low predictions near site I with current emissions on day 3, such that the peak 8-hour daily maximum ozone concentration predicted for that day is 65 ppb (rather than the 91 ppb shown in Figure 3.2).

Find: The relative reduction factor near site I ($(RRF)_I$).

Solution: (1) Calculate the mean peak 8-hour daily maximum ozone concentration obtained near site I for current and future emissions. Exclude results for day 3 from the calculations. From Figure 3.2,

(a) $(\text{mean peak 8-hr daily max})_{\text{future}} = (87 + 82 + 81)/3 = 83$ ppb

(b) $(\text{mean peak 8-hr daily max})_{\text{current}} = (98 + 100 + 90)/3 = 96$ ppb.

(2) Compute the relative reduction factor by taking the ratio of (1)(a) over (1)(b).

$(RRF)_I = 83/96 = 0.86$

Recommendations. **States should estimate current monitored design values (DVC) for each monitoring site in the nonattainment area using the procedures illustrated in Tables 3.1 and 3.2. States should consider modeled 8-hour daily maxima from all surface grid cells near a monitoring site. Default recommendations for "near" are provided in Table 3.3. Site-specific relative reduction factors (RRF) should be calculated by taking the ratio of mean highest 8-hour daily maxima obtained for the future and current emissions. A day may be excluded from consideration at a site if the nearby peak modeled current 8-hour daily maximum ozone concentration on the day is < 70 ppb.**

3.2 How Do I Apply The Recommended Modeled Attainment Test?

States should apply the modeled attainment test at each monitoring site within a nonattainment area observing a current design value of 75 ppb or greater. The test should also be applied at monitors (with current design values ≥ 75 ppb) outside the nonattainment area but affected by transport from the area. If the focus of a SIP revision is to identify local measures needed to supplement national or large scale regional measures, "sites affected by transport" are those within one day's travel time of the nonattainment area. One day's travel time may be estimated using trajectory models. If a trajectory analysis suggests that emissions occurring within the nonattainment area shortly before sunrise affect a monitor's readings shortly after sunset or earlier on the same day, States should assume that the site is "affected by transport"from the nonattainment area. For urban scale analyses (i.e., see Section 13.0), the modeled attainment test generally applies to all ozone monitors within the urban modeling domain having current monitored design values ≥ 75 ppb.

Inputs described in Section 3.1 are applied in Equation (3.1) to estimate a future design value at all monitors for which the modeled attainment test is applicable. The 8-hour NAAQS is met in an area if, over 3 consecutive years, the average 4th highest 8-hour daily maximum ozone concentration observed at each monitor is ≤ 0.08 ppm (i.e., ≤ 84 ppb using rounding conventions)[6]. Thus, if all resulting predicted future design values (DVF) are ≤ 84 ppb, the test is passed. The modeled attainment test is applied using 4 steps.

Step 1. Compute a site-specific current design value (DVC) from *observed* data.

This is done as illustrated in Tables 3.1 and 3.2. The values in the right hand column of Table 3.2 are site-specific design values.

Step 2. Use air quality modeling results to estimate the relative reduction factor for that site.

---

[6]40CFR Part 50.10, Appendix I, paragraph 2.3

This step begins by computing the mean peak 8-hour daily maximum ozone concentration for future and current emissions. This has been illustrated in Examples 3.2 and 3.3. The relative reduction factor for site I is given by Equation 3.2.

$$(RRF)_I = (\text{mean peak 8-hr daily max})_{future} / (\text{mean peak 8-hr daily max})_{current} \quad (3.2)$$

Using Equation (3.2), the relative reduction factor is calculated as shown in the column (5) in the last row of Table 3.4. Note that the RRF is calculated to two significant figures to the right of the decimal place. The last significant figure is obtained by *rounding* with values of "5" or more rounded upward. For the illustration shown in Table 3.4, we have assumed the same 4 days described previously in Example 3.3 have been simulated. We have also assumed that the monitored current design value at site I is 102 ppb.

Step 3. Multiply the observed current design value obtained in Step 1 times the relative reduction factor obtained in Step 2. If the estimated future design value is ≤ 84 ppb, the test is passed at the monitoring site being evaluated.

These calculations are illustrated in Table 3.4. Note that on day 3, the predicted current modeled peak 8-hour daily maximum ozone concentration was very low (< 70 ppb). As discussed in Section 3.1, predictions for this day are not included in calculating the mean values shown in the last row of the Table.

In Table 3.4, we see (column (2)) that the current observed design value at Monitor I is 102 ppb. Using Equation (3.1), the predicted future design value for site I is,

$$(DVF)_I = (102 \text{ ppb})(0.86) = 88 \text{ ppb}$$

Note that in this example the modeled attainment test is not passed at monitor I.

Step 4. Repeat steps 1-3 for all ozone monitoring sites with current design values ≥ 75 ppb during the 3 years used as the basis for the current monitored design value. If the test is passed at each site, the modeled attainment test is passed.

Table 3.4. Example Calculation of a Site-Specific Future Design Value $(DVF)_I$

(1) Day	(2) Observed Current Site-specific Design Value, $(DVC)_I$, ppb	(3) Current Peak Predicted 8-hr Daily max.conc near Monitor, ppb	(4) Future Peak Predicted 8-hr Daily max.conc near Monitor, ppb	(5) Relative Red. Factor (RRF), Col.(4) / Col.(5)	(6) Future Site-Specific Design Value, $(DVF)_I$, ppb Col.(5) x Col.(2)
1	102	98	87	-	-
2	102	100	82	-	-
3	102	65 (Not Considered)	Not Considered	-	-
4	102	90	81	-	-
Mean	102	96	83	0.86 (i.e., 83/96)	88

Recommendations. The modeled attainment test is applied in 4 steps.

1. Compute a site-specific current design value from *observed* data.

2. Use air quality modeling results to estimate a site-specific relative reduction factor.

3. Multiply the relative reduction factor obtained in step 2 times the site-specific monitored design value obtained in step 1. The result is an estimated site-specific future design value. If this value is $\leq$ 84 ppb, the test is passed at the monitor site being evaluated.

4. Repeat steps 1-3 for all ozone monitoring sites with current monitored design values $\geq$ 75 ppb during the 3 years used to compute the current monitored design value. If the test is passed at each site, the modeled attainment test is passed.

3.3 What Are Key Implications Of The Recommended Modeled Attainment Test?

The recommended modeled attainment test for the 8-hr ozone NAAQS raises some implications which warrant further discussion.

1. The attainment test focuses on monitoring sites. In this sense, the modeled attainment test is identical to the monitored test, used to define whether or not attainment occurs. One shortcoming of this approach is that there are usually only a small number of monitoring sites compared to the area which is modeled. This could result in the model predicting 8-hour daily maximum concentrations in one or more locations which are higher than any predicted near a monitor. If this result occurs consistently (e.g., predicted current 8-hour daily maxima at a location is > 5% above any predicted near monitored locations on 50% or more of the days modeled), it should be investigated further. As described in Section 3.4, a screening test should be completed if there is one or more locations with predictions which are consistently higher than any near a monitoring site.

If the attainment test and screening test are passed or close to being passed (e.g., the predicted future design value is < 90 ppb), a State may choose to perform a weight of evidence analysis to assess whether attainment is likely. As discussed in Section 4.0, a weight of evidence determination includes an assessment of other model outputs which are less dependent on monitored information. Other important considerations in the weight of evidence assessment would be evidence that the State plans to deploy an ozone monitoring site in a location(s) where the modeling suggests there could be future problems meeting the NAAQS and is planning to perform a subsequent review to refine its selected control strategy, if necessary. The weight of evidence approach for demonstrating attainment is discussed in Section 4.0.

2. The attainment test makes assumptions about the shape of a distribution of average 8-hour daily maxima after controls are in place. What is the basis for these assumptions? The attainment test assumes that the distribution of 8-hour daily maxima for each site will flatten after controls are established (e.g., the difference between the 90th percentile concentration and the median concentration diminishes after controls are in place), but that a day's ranking in the tail of a precontrol distribution is similar to its ranking in the post-control distribution. The first of these assumptions is common sense. Concentrations which are close to background levels already are unlikely to be greatly affected by control measures (Lefohn, et al., 1998). This assertion is also borne out by results of a study by Pacific Environmental Services (PES), Inc. (1997) showing trends in observed distributions of 8-hour daily maxima in 9 cities between 1980 and 1995. The PES study also shows that relative change observed in typical 95th and 99th percentile 8-hour daily maximum ozone observations between 1983 and 1995 is generally similar. A modeling study by Meyer, et al. (1997) lends further support by showing for numerous cities that if a day is ranked as a severe one in the pre-control distribution of predicted 8-hour daily maxima, it will have a similar ranking in the predicted post-control distribution. Our previously identified finding that a relative reduction factor tends to increase when current predicted 8-hour daily maximum ozone is below about 70 ppb but is independent at higher

predicted current values is consistent with these other studies.

3. The test adjusts observed concentrations during a current period (e.g., 1997-1999) to a future period (e.g., 2010) using model-derived "relative reduction factors". It is important that emissions for the base period used in the test correspond with the period reflected by the current design value (e.g., 1997-1999). Large deviations from this constraint may diminish credibility of the relative reduction factors. For example, if 1990 emissions were compared to emissions projected for 2010, the contrast would probably be greater than if 1996 or 1999 emissions were compared to the 2010 estimates. Presumably, this would lead to larger predicted relative differences in ozone between 1990 and 2010 vs. between 1996 or 1999 and 2010. If the resulting smaller relative reduction factors were applied to 1997-1999 observed design values, a State may underestimate 2010 ozone concentrations.

Unfortunately, the constraint described in the preceding paragraph may introduce confusion about just what we mean by "emissions for the base period". That is, the term "base case" is commonly understood to mean emissions corresponding to the episode we are modeling. For example, if we were modeling a 1991 episode, "base case" emissions would be 1991 emissions. As described in Section 16.0, it is essential to use "base case" emissions *together with* meteorology occurring in the modeled episode in order to evaluate model performance. However, once the model has been shown to perform adequately, it is no longer necessary to model the "base case" emissions. It now becomes important to model emissions corresponding to the period with the current observed design value (e.g., 1997-1999) and the period corresponding to the statutory attainment date (e.g., 2010). In this guidance, we refer to the former as the "current" emissions, and continue to refer to emissions used for model performance evaluation as "base case" emissions.

It is desirable to model meteorological episodes occurring during the period reflected by the current design value (e.g., 1997-1999). This avoids the need to derive a "current" inventory which differs from the "base case" inventory. It also avoids the need for an additional air quality model simulation corresponding to a "current" inventory which differs from the "base case" inventory. However, episodes need not be selected from the period corresponding to the current design value, provided they are representative of meteorological conditions which commonly occur when exceedances of 0.08 ppm occur. The idea is to use selected representative episodes to capture sensitivity of predicted ozone to changes in emissions during such commonly occurring conditions. There are at least two reasons why using episodes outside the period with the current design value may be acceptable: (1) availability of air quality and meteorological data from an intensive field study, and (2) availability of a past modeling analysis in which the model performed well and with which the State is satisfied.

Recommendations. States should review absolute projected future model predictions for 8-hour daily maxima to identify locations with consistently higher predictions than any near a monitoring site. An additional screening test is needed for identified locations. To apply the recommended modeled attainment test, States should use

emission estimates which correspond (a) with the period represented by the current design value (e.g., 1997-1999), and (b) with a year two years prior to the required attainment date.

3.4 What Is A Screening Test, And Why Is It Needed?

An additional review is necessary, particularly in nonattainment areas where the ozone monitoring network just meets or minimally exceeds the size of the network required to report data to AIRS. This review is intended to ensure that a control strategy leads to reductions in ozone at other locations which could have current design values exceeding the NAAQS were a monitor deployed there. The test is called a "screening" test, because if a current design value were measured at a location identified in the test, modeled results suggest it *might* exceed any at sites with available measurements.

The additional review is in the form of a screening test which should: (1) identify areas in the domain where absolute predicted 8-hour daily maximum ozone concentrations are *consistently* greater than any predicted in the vicinity of a monitoring site, and (2) for each identified area, multiply a location-specific relative reduction factor times the areawide current design value for the nonattainment area to estimate a "future design value". If the resulting estimates are less than or equal to 84 ppb at all flagged locations, the screening test is passed.

In the first part of the screening test, the word "consistently" is important. An occasional prediction which exceeds any near a monitor is not necessarily indicative of violating a NAAQS which focuses on the 4th highest daily maximum concentration, averaged over 3 consecutive years. Interpretation of "consistently" is discretionary for those implementing the modeling protocol. However, in the absence of any stronger rationale, we recommend the following default criterion:

> -predicted 8-hour daily maxima at the location in question is > 5% higher than any near a monitored location on 50% or more of the modeled days.

The "5%" difference is consistent with the size of rounding differences at 0.08 ppm. Occurrence of such a difference on 50% or more of the modeled days increases the likelihood that a difference might show up in a design value averaging observations over 3 years should a monitor be deployed at the flagged location.

What do we mean by "areas in the domain" in the first part of the screening test? For each modeled day, States should consider individual surface grid cells with predictions more than 5% greater than any "near" any monitoring site. An array of cells, centered on the identified cell, should be considered "near" the monitor (see Table 3.3). As a result, several cells may be identified for each modeled day. If any surface cell shows up within these arrays on 50% or more

of the modeled days, a future design value should be estimated for that cell using screening procedures described in the following paragraphs.

Once one or more locations is identified with current predictions consistently exceeding those near any monitor, we recommend applying a screening method to estimate future design values for such locations. The screening method applies an equation similar to Equation (3.1).

For location j,

$$(DVF_{est})_j = (RRF)_j\ (DVC_{areawide}) \qquad (3.3)$$

where

DVF_{est} = the estimated future design value obtained with the screening Equation (3.3), ppb;

$(RRF)_j$ = the relative reduction factor at location j, computed as recommended in Section 3.1, unitless.

$(DVC_{areawide})$ = the design value for the nonattainment area, ppb.
This is the highest of the site-specific monitored design values. Thus, for the example shown in Table 3.2, $(DVC_{areawide})$ would be 94 ppb.

<u>Recommendations</u>. **In addition to applying the modeled attainment test, States should apply a screening test to calculate a future design value at all locations with consistently higher current predicted 8-hour daily maxima than those near any monitoring site. The screening test is passed if the predicted future design value is ≤ 84 ppb at all flagged locations.**

4.0 If I Use A Weight Of Evidence Determination, What Does This Entail?

In Section 9.0, we note that it is preferable for States to analyze available emissions, meteorological and air quality data to provide insights concerning appropriate inputs and assumptions to include in air quality modeling analyses. States may wish to use outcomes of corroboratory analyses to provide support for results of a modeled attainment or screening test, or to conclude that attainment is (un)likely despite the outcome of the tests. In Section 4.0, we describe a weight of evidence determination, note key corroboratory analyses which should be included, and identify criteria which should be met to use additional corroboratory analyses in a weight of evidence determination.

A weight of evidence determination examines results from a diverse set of analyses, including outcomes of the attainment and screening tests. Each type of analysis has an identified outcome or set of outcomes consistent with a hypothesis that a proposed control strategy is sufficient to meet the NAAQS within the required time frame. If such an outcome occurs, then results of that analysis support the hypothesis that the proposed strategy is adequate. Each analysis is weighed qualitatively, depending on the ability of the analysis to address adequacy of a strategy and on the credibility of the analysis. If most (i.e., overall weight of) evidence produced by the diverse analyses supports the hypothesis, then attainment of the NAAQS is demonstrated with the proposed strategy. The end product of a weight of evidence determination is a document which describes analyses performed, data bases used, key assumptions and outcomes of each analysis, and why a State believes that the evidence, viewed as a whole, supports a conclusion that the area will attain the NAAQS. The further a modeled attainment and screening test are from being just passed, the more convincing other evidence needs to be to reach a different conclusion in a weight of evidence determination. As noted in Table 2.1, if a modeled attainment or screening test predicts one or more future design values $\geq$ 90 ppb, we believe it is doubtful that more qualitative arguments based on other analyses can be sufficiently convincing to support a conclusion that attainment will occur.

Each weight of evidence determination will be subject to area-specific conditions and data availability. Area-specific factors may affect the types of analyses which are feasible for a nonattainment area, as well as the significance of each. Thus, decisions concerning which analyses to perform and how much credence to give each needs to be done on a case by case basis by those implementing the modeling/analysis protocol. In Section 4.1, we identify several recommended core analyses which should be used to corroborate one another in a weight of evidence determination. It is appropriate to require considering a core set of analyses to reassure reviewers and the public that a selective set of analyses, supporting a particular viewpoint, has not been chosen. If it is not feasible to perform one or more of the core analyses, a State should document why not. In Section 4.2, we note that additional, optional corroborative analyses may be performed. We provide several examples of such analyses, and identify conditions which should be met for them to be considered in a weight of evidence determination.

4.1 What Analyses Should I Consider In A Weight Of Evidence Determination?

At a minimum, a weight of evidence determination should consider the following 3 types of corroborative analyses: application of air quality models, observed air quality trends and estimated emissions trends, and outcome of observational models.

Table 4.1 addresses each of the 3 recommended core analyses. In the table, we identify factors which might cause those implementing the modeling/analysis protocol to give greater credence to a particular set of results (column (2)). We also identify outcomes for each analysis consistent with a hypothesis that emission reductions implied by a strategy are adequate to demonstrate attainment (column (3)).

We discuss each of the recommended core corroboratory analyses in the following subsections.

4.1.1 Air Quality Models

Weight given to results obtained with air quality models depends on how good the model performance is as well as the rigor with which the performance has been tested. Figure 3.3 suggests that a relative reduction factor averaged over as many as 10 days is insensitive to the mean 8-hour daily maximum ozone concentration predicted for these days. However, for practical reasons, it may not be feasible to use 10 or more days at every site in the modeled attainment test. It is conceivable that the relationship between RRF and current predicted 8-hour daily maxima could become noisier with smaller sample sizes. Thus, model results have higher credibility if nearby maximum predicted 8-hour daily maxima agree with observed values at the monitoring sites. It is not possible to offer definitive guidance on how close the agreement should be. However, we suggest that model results may have higher credibility if nearby predicted maximum 8-hour daily maxima agree within about 20% of observations on most or all of the days used to compute the relative reduction factors near each monitor. This "20% difference" is roughly equivalent to the difference between the design value concentration sufficient to pass the modeled attainment test (i.e., 84 ppb) and the concentration at which RRF begin to become sensitive to current predicted 8-hour daily maxima (i.e., < ~70 ppb). Performance evaluation is discussed further in Section 16.0.

Model applications for which an extensive observational data base exists have greater credence, especially if the data base includes monitored values of indicator species and precursor data. For ozone, one of the most uncertain inputs to a modeling analysis is the emission projections which must be made to a future year(s) of interest. This uncertainty is reduced if the projection period is short. Hence, weight of evidence provided by modeling is increased with short projection periods. If rigorous quality assurance and review is provided for the model's emissions and meteorological inputs, this may increase confidence that the model is yielding correct answers for the right reasons. Thus, rigor used in preparing model inputs also increases

Table 4.1. Recommended Core Analyses for a Weight of Evidence Determination, Factors Affecting Their Credibility and Outcomes Consistent with Meeting the NAAQS

(1) Type of Analysis	(2) Factors Increasing Credibility of the Analysis	(3) Outcomes Consistent with Hypothesis That a Candidate Strategy will Lead to Attainment
Air Quality Models	-good model performance -extensive observational data base available -short projection periods -carefully quality assured emissions inventory -confidence in meteorological inputs -good ability to pose and address questions about a strategy's adequacy -other analyses tend to corroborate conclusions -selected episode days have observations near the design value	-the modeled attainment test is passed -screened estimates for future design values at sites w/o monitors are < 84 ppb -the attainment/screening tests are nearly passed, the control strategy requires additional reductions and efforts are underway to subsequently review/refine the strategy -commitment is made to deploy monitors at locations not passing the screening test -substantial modeled improvement in air quality is predicted using several measures described in Section 4.1.1 -similar conclusions are reached with other peer reviewed models

Table 4.1. Recommended Core Analyses for a Weight of Evidence Determination, Factors Affecting Their Credibility and Outcomes Consistent with Meeting the NAAQS (continued)

(1) Type of Analysis	(2) Factors Increasing Credibility of the Analysis	(3) Outcomes Consistent with Hypothesis That a Candidate Strategy will Lead to Attainment
Analysis of Air Quality and Emissions Trends	-current or future (air quality model) predicted design value is within a few ppb above 84 ppb -extensive monitoring network exists -both ozone and precursor trends are available -statistical model used to normalize trend for meteorological differences explains much variance -short projection periods used in the analysis -a pronounced, statistically significant downward trend is apparent -similar conclusions are reached using several trend parameters -continued, comparable relative reductions in emissions are provided for	-a pronounced downward normalized trend exists in the site-specific design value at all sites with design values greater than 84 ppb. -Using projected emissions to extrapolate the air quality trend line to the required attainment date indicates an 8-hour daily maximum concentration ≤ 84 ppb. -Other observed air quality trend parameters also show a substantial improvement.

Table 4.1. Recommended Core Analyses for a Weight of Evidence Determination, Factors Affecting Their Credibility and Outcomes Consistent with Meeting the NAAQS (concluded)

(1) Type of Analysis	(2) Factors Increasing Credibility of the Analysis	(3) Outcomes Consistent with Hypothesis That a Candidate Strategy will Lead to Attainment
Use of Observational Models	**-an extensive monitoring network exists** **-precursor and indicator species are measured using instruments with appropriate sensitivity** **-monitoring sites appear spatially representative** **-data have been quality assured, and results are self-consistent** **-plausible physical explanations exist for findings**	**-Findings indicate sources controlled in the candidate strategy are important causes of observed high ozone.** **-Analysis of indicator species suggests the direction of the strategy (e.g., emphasis on VOC or NOx) is appropriate.**

credibility given to the results. Considering days with concentrations near site-specific design values increases confidence that relative reduction factors, developed for use in the tests, are appropriate.

Selecting episode days which have occurred recently is also advantageous. This follows if there is a long term downward trend in observed ozone. Selection of a "severe" episode (compared to recent observations) from an earlier year may not actually be so severe. In Section 12.0, we suggest a method for characterizing severity which circumvents this problem.

Proper selection of episode days increases confidence in the results of a modeled attainment test. Confidence in the quantitative results of a modeled attainment test is greater if corroborative, more qualitative, analyses yield supporting conclusions about appropriateness of a strategy. Finally, of the analyses available, modeling reflects the most comprehensive attempt to integrate emissions and meteorological information with atmospheric chemistry. As such, modeling has the greatest capability to address questions about adequacy of a strategy to meet air quality goals in the future. Thus, States should include modeling results in a weight of evidence determination, and these should ordinarily be very influential in deciding if attainment will occur.

The outcome from modeling supports use of a proposed strategy for attainment if the modeled attainment and screening tests, described in Section 3.0, are passed. If the screening test for future design values at locations without monitors is not passed, a commitment to deploy

ozone monitors at such locations should be an important consideration in approving an attainment demonstration. In general, the closer modeled output is to passing the attainment and screening tests, the easier it is for other analyses to produce evidence which supports attainment. If a modeled attainment or screening test is not passed, selection of a strategy which substantially reduces precursor emissions and an agreement to perform a subsequent review (based on improved data bases/tools) to refine the strategy, if necessary, can be considered in a weight of evidence determination if other modeling outputs and other analyses support a conclusion that the current selected strategy may lead to attainment. Other model-produced indicators that a proposed strategy may be adequate are (1) model predictions show major improvements in air quality using a variety of measures, and (2) other peer reviewed atmospheric simulation models suggest attainment occurs using modeled attainment test results or other outputs.

We recommend that at least 3 additional model outputs be examined in weight of evidence determinations to provide assurance that passing or nearly passing the recommended modeled attainment and screening tests indicates attainment. Like the tests, and for similar reasons, each of these additional outputs reflects *relative* changes in predicted air quality. States may use other model outputs (not described herein) in a weight of evidence determination as well.

1. Compute the relative change in surface grid-hours > 84 ppb in the nonattainment area.

This output reflects the frequency with which predicted *hourly* concentrations exceed the concentration specified in the 8-hour NAAQS. Such a measure is not directly related to the form of the NAAQS. Further, if current and future predictions are subject to a systematic bias, this output could be misleading. However, if modeled episodes are chosen to represent a variety of meteorological conditions under which the NAAQS is exceeded at one or more monitoring sites, and the model performs well using the measures described in Section 16.0, a large reduction in the frequency of predicted hourly concentrations of 85 ppb or more is consistent with a conclusion that a proposed strategy would meet the NAAQS.

The decision about what constitutes a "large" reduction in the predicted frequency of hourly concentrations > 84 ppb is subjective, since it is difficult to relate this measure to the NAAQS. An example of "large" would be an "80%" reduction.

2. Compute the relative change in the number of grid cells in the nonattainment area with predicted 8-hr daily maxima > 84 ppb.

This output estimates reduction in the pervasiveness of estimated 8-hour concentrations in excess of the concentration specified in the NAAQS. It is subject to the same caveats as the preceding output. One additional complication may occur if there are not many surface grid cells in which current emissions lead to 8-hour daily maximum ozone estimates > 84 ppb. An 80% reduction in this measure may be regarded as an example of a "large" reduction.

3. Compute the relative change in the total difference (ppb-hr) of hourly predictions > 84 ppb in the nonattainment area.

Although not the same, this output is similar in concept to the change in the "dosage" to concentrations greater than 84 ppb. Since we are interested in estimating the likelihood that a strategy will lead to attainment rather than estimating reduction in total dosage, this output should be calculated differently from the procedure ordinarily used to calculate dosage metrics, like SUM06. We recommend using Equation (4.1).

$$RD = \frac{\sum_{j}^{G}\sum_{i}^{N}(C_{ij}-84)_{future}}{\sum_{j}^{G}\sum_{i}^{N}(C_{ij}-84)_{current}} \qquad (4.1)$$

Where

RD = Relative Difference
C_{ij} = Predicted 1-hour concentration for hour I and grid cell j greater than 84
N = Total number of hours
G = Total number of grid cells

This metric is subject to the same caveats as the preceding two metrics. As with the other two we suggest an 80% reduction (i.e., leading to a value of ≤ 0.20 in equation (4.1)) as an example of a "large" reduction.

4.1.2 Analysis Of Air Quality And Emissions Trends

This approach is to normalize air quality trends observed over a period for meteorological differences occurring from year to year. The Cox/Chu approach, used extensively in U.S.EPA (1996), is one example of how air quality trends can be normalized (Cox, et al., 1993, 1996). Other procedures can also be used. A curve is fit through the normalized trend and extrapolated to the year in which the air quality goal is to be met. Extrapolations are made by considering past trends as well as past and projected emission reductions. If the trend is statistically significant, the extrapolated value for the attainment year is at or below the air quality goal, and projected relative emission reductions are comparable or greater than reductions occurring during the period for which the trend is constructed, results of the trend analysis suggest a strategy will be adequate. This procedure is illustrated in Figure 4.1 and by the following example.

Example. Estimate the relative reduction in emissions (VOC, NOx or both) occurring during the period corresponding to the observed normalized trend in the average 4th highest 8-hour daily maximum in the nonattainment area. Use the estimated emission trend in concert with the normalized air quality trend to determine an "emission reduction sensitivity factor" (e.g., (ppb)/(percent emission reduction)). Multiply the sensitivity factor times the percent reduction in emissions projected between the current period and the required attainment date. Subtract the result from the current design value to get a projected design value. If the projected design value is $\leq$ 84 ppb, the trend analysis supports a hypothesis that a proposed control strategy will suffice to reach attainment by the required date.

The trend analysis we suggest assumes that a linear extrapolation of observed past correspondence between monitored design values and estimated emission changes will accurately describe future air quality. This assumption probably works best if the current design value nearly meets the goal of being $\leq$ 84 ppb. Weight given to trend analyses depends on several other factors as well. The more air quality data available and the greater variety of trend parameters which show major improvements, the more credible the results. In the case of ozone, availability of trends in ambient precursor data which are consistent with the emissions and ozone trends also lend credibility to the results. Weight of evidence produced by trend results is higher if the procedure used to normalize the trend for meteorological differences explains much of the variability attributable to these differences. Finally, trend analysis is more believable if it is not necessary to extrapolate very far into the future.

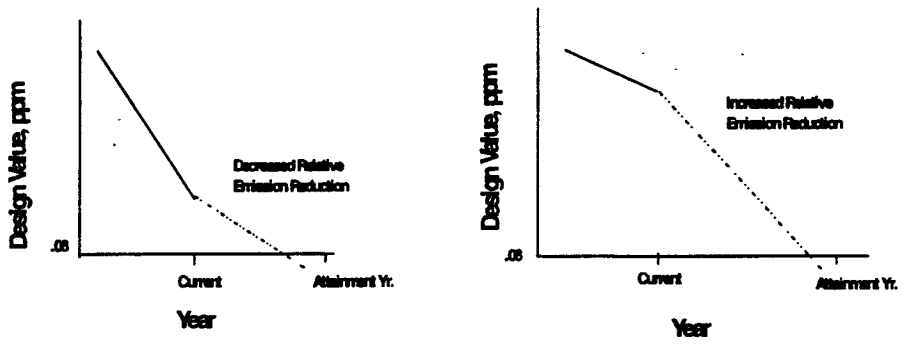
4.1.3 Use Of Observational Models

Observational models take advantage of monitored data to draw conclusions about the relative importance of different types of VOC and/or NOx emissions as factors contributing to observed ozone. There are at least 4 approaches with potential for doing this: receptor models (Watson (1997), Henry, et al. (1994) and Henry (1997, 1997a, 1997b)), indicator species approach (Sillman (1997, 1998)), smog produced algorithm approach (Blanchard, et al., (1997, 1999)) and a relative incremental reactivity approach (Cardelino, et al., (1995)).

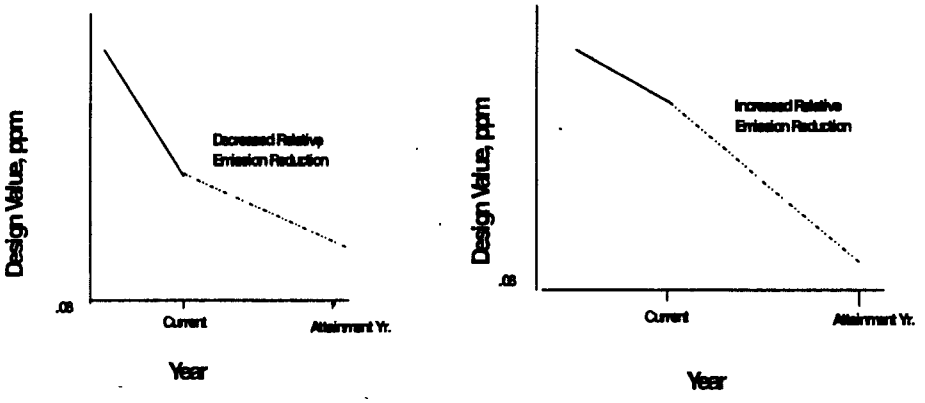
Observational models are potentially useful for assessing whether a proposed strategy is oriented toward source categories whose emissions appear to be associated with current observed high ozone. However, their ability to estimate *how much* control is needed is limited, unless one can justify assuming an approximately linear relationship between precursor emissions and observed 8-hour daily maximum ozone. Thus, observational approaches are ideally suited to corroborate results obtained for ozone with more quantitative techniques, like air quality models. Like air quality models, observational models are also subject to uncertainties. Thus, results which are ambiguous should not be considered at odds with conclusions reached with an air quality model.

Figure 4.1. Examples Showing Use of Trend Analysis in Attainment Demonstrations

(a) Cases suggesting attainment is likely



(b) Cases yielding results inconsistent with attainment



Observational models which rely on use of indicator species can be used to show whether or not ozone may be sensitive to the types of precursors (i.e., VOC or NOx) reduced by a particular control strategy. Receptor models, like the chemical mass balance approach, may be useful for confirming whether a strategy is reducing the right sorts of sources. Observational models can be used to examine days which have not been modeled with an air quality model, as well as days which have been modeled. The resulting information may be useful for drawing conclusions about the representativeness of the responses simulated with the air quality model for a limited sample of days.

Summarizing, if conclusions drawn with one or more observational models suggest that the types of sources to be controlled under a proposed strategy are those that appear associated with high ozone and/or are those to which observed ozone is sensitive, this supports a hypothesis that the strategy is directionally correct.

Strength of the evidence produced by observational models is increased if an extensive monitoring network exists and at least some of the monitors in the network are capable of measuring pollutants to the degree of sensitivity required by the methods. Evidence produced by observational models is more compelling if several techniques are used which complement one another and produce results for which plausible physical/chemical explanations can be developed. Indications of a strong quality assurance analysis of collected data and measurements that are made by a well trained staff also lend credence to the results.

Recommendations. **Weight of evidence determinations are best performed using a variety of diverse analyses in a corroborative fashion. Prior to its application, each selected analysis should have identified outcomes consistent with concluding a proposed strategy is adequate. At a minimum, States should consider the following 3 types of corroboratory analyses in a weight of evidence determination: (1) output from air quality model(s) (i.e., modeled attainment test results plus other indicators), (2) air quality and emission trend analysis, and (3) interpretation of results obtained with observational models. If it is not feasible to include one or more of the recommended corroboratory analyses, the reasons why not should be documented.**

4.2 What If I Want To Consider Additional Corroborative Analyses?

The list of analyses in Section 4.1 is not an exhaustive one. A State may use other types of analyses to supplement the core set recommended for a weight of evidence determination. To have another type of analysis considered, a State should identify why it believes the analysis will produce information which has a bearing on attainment of the NAAQS. In addition, the procedure to be used in applying the method and the extent of the data base available to support it should be identified. Finally, prior to application of the method, a State should identify outcomes which would be consistent with a hypothesis that a proposed emission reduction strategy will lead to attainment.

Identity of additional corroborative analyses is, in part, a function of the available data base and analytical tools, as well as questions posed by the outcomes of the recommended core corroboratory analyses. For purposes of illustration, we identify some additional analyses of the sort which States might consider. None of these is required, and States may well choose to consider other optional analyses or no optional analyses at all.

Quantifying uncertainty associated with air quality model estimates. In this guidance, we recommend that "uncertainty" be accounted for using a modeled attainment test which uses models in a "relative" sense and by recognizing that use of corroboratory analyses may be desirable in a weight of evidence determination. Thus, we account for uncertainty in a qualitative way, without actually estimating it.

States may find it useful to quantify estimates of uncertainty and then use these results qualitatively in a weight of evidence determination. In Section 6.0, we identify three diagnostic tests which may be useful for this purpose. The first of these is one which has been proposed by Reynolds, et al., (1997). This test is to prepare "alternative base case" emission estimates, reflecting reasonable alternative assumptions about current emissions which lead to comparable or better model performance. Note differences in projected design values from these alternative current emissions. A second test is to assume alternative (reasonable) growth assumptions. This could reflect using differing growth rates or placement of new sources in different, equally probable, locations. Note the differences in projected design values for the different growth assumptions. Combinations of the first two tests are also possible. A third test is one in which a control strategy under serious consideration is simulated with an alternative grid resolution or with different (reasonable) meteorological assumptions. For example, due to resource constraints, it might be necessary to initially select a strategy using a grid with 12 km grid cells (horizontal dimension). Differences in projected air quality obtained with a grid having 4 km cells could then be ascertained.

Other approaches for estimating uncertainty have been described in the literature (Gao, et al., (1996) and Yang, et al., (1995)). Many of these approaches also assess sensitivity of model predictions to uncertainties in input variables. For outcomes to be most relevant to the way we recommend models be applied in attainment demonstrations, it is preferable that such procedures focus on sensitivity of estimated relative reduction factors (RRF) and resulting projected design values to the variations in inputs or model formulations.

Once a range in projected design values is obtained using tests like those previously described, a qualitative assessment can be made of how likely it is that a strategy will lead to attainment of the NAAQS. For example, if most of the results lead to projected design values $\leq$ 84 ppb, this supports a conclusion that the strategy, if implemented, will demonstrate attainment. Choice of tests and interpretation of the outputs should be agreed upon beforehand in concert with the appropriate U.S. EPA Regional Office.

Compare monitored design values for the current period used in the test with those measured in other periods. The objective of this analysis is to assess whether current design values used in the modeled attainment and screening tests are atypically high or low due to natural or meteorological conditions. If the current design values are lower (higher) than normal, the tests would yield overly optimistic (pessimistic) results.

An analysis of current design values is complicated by trends in emissions. For example, one would expect design values measured several years previously to be higher than current values if there has been an ongoing program to reduce precursor emissions. This does not necessarily imply that the current design value is atypically low. The problem of emissions trends can be addressed by examining statistical relationships between meteorological conditions and 8-hour daily maximum concentrations to see whether conditions corresponding to high ozone occurred more or less frequently than usual during the current period. Approaches which are analogous to that described by Cox, et al., (1996) or by Deuel, et al., (1998) might be tried to see how the current 3-year period ranks with other 3-year periods in terms of its meteorological ozone forming potential.

States may use results of an assessment of the severity of the current monitored areawide design value qualitatively to see if the value is atypical (e.g., in the lower or upper quartile). If the current monitored areawide design value as judged to be atypically high (low), this could be used to support an argument that the control target implied by the modeled attainment test is too restrictive (not restrictive enough).

Examine Basis for including/excluding days from calculations made in the modeled attainment test. This analysis would examine *observed* air quality and meteorological conditions to refine the basis for previously calculated relative reduction factors. That is, has a modeled day been improperly included or excluded from the calculations at a particular monitoring site? Recall the intent of the attainment test is to consider days with meteorological conditions which correspond to periods when observed 8-hour daily maximum ozone concentrations are close to or exceed 84 ppb. In Section 3.1, we suggest using model predictions as the basis for excluding improper days (i.e., days with already low ozone) from the calculations. It may be possible to refine choice of days used for a site using available air quality and meteorological data. For example, States may examine days used to calculate the relative reduction factor to ensure they reflect wind orientations corresponding with observed concentrations exceeding 84 ppb.

Recommendations. Optional analyses may be considered in addition to the 3 recommended analyses identified in Section 4.1. To use an optional analysis in a weight of evidence determination, a State should (1) explain the rationale for the analysis, (2) identify the data base underlying the analysis, (3) describe the methodology to be used in applying the analysis, and (4) identify outcomes which would be consistent with a hypothesis that a proposed control strategy will suffice to attain the NAAQS.

5.0 How Can I Improve Modeling And Other Analyses In Weight Of Evidence Determinations?

In Section 4.0, we identified a set of analyses which should be considered when performing a weight of evidence assessment of whether a proposed control strategy will lead to attainment of the 8-hour ozone NAAQS. To be most credible, modeling and many of the other analyses rely on presence of good emissions and ambient data bases. Although commitment to undertake subsequent review of a SIP revision is not a prerequisite for approval of the revision, States should anticipate a need at the required time of attainment to confirm that the NAAQS has indeed been met or to diagnose why not. In this Section, we identify measurements and activities which provide better support for modeling and other analyses in weight of evidence determinations. Resulting improved data bases may increase reliability of reviews to identify reasons for attainment or non-attainment of the NAAQS and provide a better basis for revising a control strategy, if necessary. We conclude by identifying anticipated future analyses which could benefit from prior efforts to improve available data bases.

5.1 What Data Gathering Or Other Efforts Might Be Helpful To Support Current Analyses Or Subsequent Reviews?

Efforts to improve the monitored air quality/meteorological data bases and to update and improve emission inventory estimates should lead to improved weight of evidence analyses (i.e., including modeling) and improved subsequent reviews. In this subsection, we identify types of monitoring which may prove helpful. We then briefly discuss efforts to improve the inventory. We conclude by identifying ways in which this improved information might ultimately be used.

Deploying additional air quality monitors. One type of additional monitoring which should be considered has already been mentioned in Sections 3.0 and 4.0. This is to deploy additional ozone monitors in locations which a screening test, described in Section 3.4, suggests may have future design values $\geq$ 85 ppb. This would allow a better future assessment of whether the NAAQS is being met at locations where the model now consistently predicts concentrations higher than any near existing monitoring sites.

Measurement of "indicator species" is a potentially useful means for assessing which precursor category (VOC or NOx) limits further production of ozone near the monitor's location at various times of day and under various sets of meteorological conditions (some of which may not have been previously considered with an air quality model). Sillman (1998) and Blanchard, et al., (1997, 1999) identify several sets of indicator species which can be compared to suggest whether monitored ozone is limited by availability of VOC or NOx. Comparisons are done by looking at ratios of these species. The following appear to be the most feasible for use in the field by a regulatory agency: O_3/NOy, O_3/(NOy - NOx) and O_3/HNO_3. Generally, high values for the ratios suggest ozone is limited by availability of NOx emissions. Low values suggest availability of organic radicals (e.g., attributable to VOC emissions) may be the limiting factor. For these ratios to be most useful, instruments should be capable of measuring NOy, NOx, NO_2

and/or HNO_3 with high precision (i.e., greater than that often possible with frequently used "routine" NOx measurements). Thus, realizing the potential of the "indicator species method" as a tool for model performance evaluation and for diagnosing why observed ozone concentrations do or do not meet previous expectations may depend on deploying additional monitors. States should consult the Sillman (1998) and Blanchard, et al. (1997, 1999) references for further details on measurement requirements and interpretation of observed indicator ratios.

Receptor models are another class of observational approaches which is potentially useful for corroborating assumptions made in air quality models or for diagnosing reasons for unexpected air quality observations in a subsequent review of for use in ozone-related applications, receptor models require observations of VOC species, such as those made in the PAMS network. Receptor models work by noting a combination of speciated source profiles which best explains speciated air quality observations on a day (chemical mass balance approach) or by noting a limited number of species which track each other well from day to day (multi variate statistical approach). Both approaches are limited by collinearity of many of the VOC species. This prevents many distinctive source categories from being identified or leads to inconclusive results concerning which source categories are contributing to observed air quality. Measuring more species is a potential means for reducing the collinearity limitation. An opportunity for doing this may exist as a result of the U.S. EPA's implementation plan for the $PM_{2.5}$ monitoring program (U.S. EPA, 1998b). This plan provides resources for measuring PM species (including some organic particulates) at approximately 300 locations. A State could collocate monitors collecting gaseous organic species at some of these sites. Availability of sites with collocated gas and aerosol phase organic measurements could increase the power of receptor models as diagnostic tools for explaining observations in subsequent reviews. In Section 5.2, we note that "organic carbon" is one of the key components of $PM_{2.5}$. This component is not so well understood as several of the others. This is true, in part, because of the volatility of some species of organic particulate. Well designed studies which measure both gaseous and aerosol phase organics at a site may lead to a better understanding of sources of organic particulates as well as ozone.

Making measurements aloft. Almost all measured ambient air quality and meteorological data are collected within 20 meters of the earth's surface. However, the modeling domain generally extends several kilometers above the surface. Further, during certain times of day (e.g., at night) surface measurements are not representative of air quality or meteorological conditions aloft. Concentrations aloft can have marked effects when they are mixed with ground-level emissions during daytime. Thus, weight given to modeling results can be increased if good agreement is shown with air quality measurements aloft. The most important of these measurements are ozone, NOy, NO, NO_2, as well as several relatively stable species like CO and selected VOC species. Measurements of SO_2 may also be helpful for identifying presence of plumes from large combustion sources. Highest priority should be given to making measurements near sunrise as well as during midday.

Measurements of altitude, temperature, water vapor, winds and pressure are also useful.

Continuous wind measurements, made aloft in several locations, are especially important. They provide a data base to "nudge" windfields, initially calculated with dynamic meteorological models, so that these estimates are more consistent with observations. This provides greater assurance that the air quality model correctly reflects the configuration of sources contributing to ozone formation. Temperature, pressure and water vapor measurements aloft provide a basis for assuring that the air quality model accurately reflects vertical exchange and mixing within the planetary boundary layer. This is a key factor affecting dilution of emissions, as well as atmospheric chemistry.

Collecting locally applicable speciated emissions data. While the U.S. EPA maintains a library of default VOC emissions species profiles (U.S. EPA, (1993)) and data at *www.epa.gov/ttn/chief/software.html#speciate*, some of these may be dated or may not properly reflect local sources. Use of speciated emissions data is a critical input to the chemical mass balance receptor model as well as to air quality models. Efforts to improve speciation profiles for local sources should thus enhance credibility of several of the procedures recommended for use in a weight of evidence determination.

Projecting emission estimates and comparing these to subsequent emission estimates. States addressing traditional nonattainment areas with lengthy attainment dates may find it worthwhile to project emissions to two future years and retain the resulting data files for use in two subsequent reviews. The first of these is the year by which attainment is the goal (e.g., 2010 for ozone). The second is some intermediate year (e.g., 2005-2007). This intermediate projection could be useful to help diagnose reasons for subsequent observed ozone trends which are inconsistent with earlier expectations obtained with an air quality model. Retention of projected emission data bases would enable States to compare the projected inventory estimates with an inventory which is subsequently updated. These checks would be possible after the inventory update for 2005 becomes available, shortly after that year. Similar comparisons would be possible for projections made in nonattainment areas which have 2005-2007 as the required attainment date.

In Section 5.2, we note it will probably be necessary to consider emissions and air quality impacts in each of the 4 seasons when addressing strategies to meet goals for $PM_{2.5}$ and visibility. Anticipating the need to integrate ozone strategies with strategies for meeting these goals, States may wish not to focus exclusively on summertime emissions of VOC, NOx and CO for purposes of developing a strategy to meet the ozone NAAQS. Estimating emissions for each of the 4 seasons and then noting effects of proposed ozone strategies on such estimates may be a factor in choosing among alternative strategies for meeting the 8-hour ozone NAAQS.

Future diagnostic analyses using air quality models. To facilitate a subsequent review, States should retain meteorological as well as current and projected emission input files developed to support the 2003 SIP revision. When a model is applied with updated emissions estimates and/or with meteorological inputs indicative of episodes chosen in 2005-2007, several useful comparisons are possible if the old files are retained. That is, a State would be better able

to determine whether differences are explained by revised emission estimates, poor choice of meteorological episodes in the initial analysis or by changes which have occurred in the model formulation during the intervening years. Insights from such comparisons should help a State explain why changes in the strategy reflected in its 2003 SIP revision may or may not be necessary.

5.2 Why Is It Desirable To Plan For A Subsequent Review?

Commitment to undertake activities supporting subsequent review of a SIP revision is not a prerequisite for approval of the revision. Thus, "why do a subsequent review and why worry about it now?" Subsequent reviews will be needed at the time attainment of the ozone NAAQS is required for nonattainment areas. The purpose of such a review is to confirm that a NAAQS has indeed been met, or to diagnose available information to determine why not. The required attainment date for traditional nonattainment areas could be as early as 2005-2007, or might be as late as 2010.

5.2.1 Integration With Attainment Strategies For 1-Hr Ozone NAAQS

A 2010 attainment date for the 8-hour NAAQS is possible in areas classified as "severe" or "extreme" nonattainment areas for the 1-hour NAAQS for ozone. For such areas, priority is given to attaining the 1-hour NAAQS. Additional measures needed to meet the 8-hour NAAQS, may be implemented after the attainment date for the 1-hour NAAQS. Thus, the projected attainment date for areas with the most serious ozone problems occurs well after 2003, when the SIP revision may be due. Further, there is a logical time (at the required time of attainment for the 1-hr NAAQS in "severe" nonattainment areas) to review and diagnose emissions and air quality data so as to refine an initial strategy for meeting the 8-hour NAAQS. Such a review, performed at an interim date, helps to ensure that strategies for meeting the 8-hr and 1-hr NAAQS for ozone are "integrated". That is, an 8-hour strategy builds upon the consequences of a strategy for meeting the 1-hour NAAQS, which are reflected in air quality and emissions data circa 2005-2007. Thus, two subsequent reviews are desirable for nonattainment areas with later attainment dates (e.g., at the time when attainment of the 1-hour NAAQS is required in "severe" nonattainment areas, and at the time the 8-hour NAAQS must be met). For nonattainment areas with more immediate attainment dates (e.g., 2005-2007), only one subsequent review is appropriate within the time frame required for attainment. This might occur at the time attainment is required.

5.2.2 Anticipated Modeling Principles For $PM_{2.5}$ And Visibility, And Integrating Ozone Strategies With Goals For $PM_{2.5}$ And Visibility

The U.S. EPA's policy is to encourage integration of control strategies to reduce ozone with those designed later to meet NAAQS for $PM_{2.5}$ and reasonable progress goals to reduce regional haze. We believe such integration will reduce overall costs of meeting multiple air quality goals. The desire to integrate strategies meeting air quality goals for ozone, $PM_{2.5}$ and

regional haze is another reason for subsequent review of an ozone SIP revision submitted in 2003 or earlier. Much of the data base used as a basis for later $PM_{2.5}$ SIP revisions will be collected during 2000-2002. Thus, the scope of the $PM_{2.5}$ problem, if any, will not be fully known at the time modeling and other analyses must be completed to support the 2003 or earlier SIP revisions for ozone. We anticipate that SIP revisions for $PM_{2.5}$ will be due about 2007-2008---about the same time as a subsequent review of the sufficiency of the previously selected strategy to meet the 8-hour ozone NAAQS. Periodic review of strategies to improve visibility is also anticipated within this time frame.

Guidance for demonstrating attainment of $PM_{2.5}$ NAAQS and reasonable progress reducing regional haze is not yet available (i.e., as of mid-1999). During its development, this guidance will be subject to intense review. Consequently, our current ideas for addressing $PM_{2.5}$ and regional haze could change. Nevertheless, we present some modeling/analysis "principles" to help States develop data bases and capabilities for considering joint effects of control strategies for ozone, $PM_{2.5}$ and regional haze in a subsequent review of the initial SIP revision for the 8-hour NAAQS for ozone.

1. Emissions and meteorological conditions vary seasonally. Effect of a control strategy on annual $PM_{2.5}$ concentrations should be assessed by estimating effects on mean $PM_{2.5}$ for each season and using the resulting information to estimate annual impacts. Emission estimates for VOC, NOx, primary $PM_{2.5}$, sulfur dioxide and ammonia will be needed.

2. PM is a mixture of component species. Each component may be attributable to causes which differ from those for others. The modeled attainment test should separately estimate effects of a control strategy on major components of the mix. Effect of a strategy on $PM_{2.5}$ can be assessed by noting the net effect of a strategy on each major component of the mix. We may recommend the following components for separate consideration:

-mass associated with sulfates;
-mass associated with nitrates;
-mass associated with organic carbon;
-mass associated with elemental carbon;
-mass associated with all other species.

3. The recommended modeled attainment test for the annual $PM_{2.5}$ NAAQS will focus on monitoring sites with speciated data. Models will be applied in a relative sense to estimate component- and site-specific relative reduction factors. Relative reduction factors will be used with current speciated design values to estimate future design values. A weight of evidence approach will be identified as an alternative to using the modeled attainment test by itself. Because the period of record for measurements is much less than that for ozone, observational models will probably be relatively more important and trend analysis relatively less so.

4. Ambient air quality data should be reviewed to assess whether exceedances of the concentration specified in the 24-hr NAAQS for $PM_{2.5}$ is a hot spot problem significantly influenced by nearby primary emissions, or a problem which is significantly influenced by more pervasive high concentrations of secondary $PM_{2.5}$. If the problem is a hot spot problem and a model performs well predicting primary $PM_{2.5}$ from the nearby source(s), a modeling approach similar to that followed for PM_{10} may be appropriate. If the problem is more pervasive, with important contributions from secondary components of $PM_{2.5}$, or model performance predicting primary $PM_{2.5}$ is poor, a relative approach similar to the approach for the annual NAAQS is likely.

5. Visibility attenuation estimates will be obtained from estimates made for each of the previously identified major components of $PM_{2.5}$. A modeled test for reasonable progress will estimate relative reduction factors for the major components of $PM_{2.5}$. These will be used with speciated $PM_{2.5}$ concentrations representative of current days with poor visibility in a Class I area to estimate representative future speciated concentrations of $PM_{2.5}$. Current and future visibility extinction coefficients will then be estimated using procedures described in Sisler (1996). Reasonable progress will be determined by comparing estimates of current and future extinction coefficients to see whether an identified improvement goal is realized.

Recommendations. The following data gathering activities may lead to more informative weight of evidence analyses and better subsequent reviews:

-deploy ozone monitors in areas where models consistently predict ozone greater than any predicted near existing monitor sites;
-make continuous meteorological measurements aloft and air quality measurements aloft, especially during early morning hours (near sunrise), as well as during midday;
-collocate sufficiently sensitive monitors to measure NOy, NO_2, HNO_3, and NOx at selected ozone monitoring sites;
-collocate monitors to measure gaseous organic species with selected monitors in the $PM_{2.5}$ monitoring network used to estimate particulate organic species;
-improve local speciated VOC emission data bases;
-retain meteorological, current and projected emission files as well as output files used in modeling the strategy reflected in the initial SIP revision for possible future diagnostic tests with newer data bases and/or models.

A State need not include plans for a subsequent review of its strategy demonstrating attainment of the 8-hour NAAQS for ozone as part of its initial SIP revision. However, a subsequent review will be needed at the time of required attainment to ascertain whether attainment has occurred. States with a protracted attainment date for the 8-hour NAAQS may also wish to consider a subsequent review at the time the 1-hour NAAQS should be met (e.g, 2005-2007). Subsequent reviews may be helpful for "integrating" strategies to meet the 8-hour ozone NAAQS with those for meeting the 1-hour NAAQS and with those addressing air quality goals for $PM_{2.5}$ and regional haze.

6.0 What Documentation Do I Need To Support My Attainment Demonstration?

States should follow guidance on reporting requirements for attainment demonstrations in U.S. EPA (1994). The first 7 subjects in Table 6.1 are similar to those in the 1994 guidance. The 1994 guidance envisions an air quality model as the sole means for demonstrating attainment. However, the current guidance (i.e., this document) identifies a weight of evidence determination as a means for corroborating the modeled attainment test in an attainment demonstration. In addition, feedback received since the earlier guidance has emphasized the need for technical review of procedures used to identify a sufficient control strategy. Thus, we have added two additional subject areas which should be included in the documentation accompanying an attainment demonstration. These are a description of the weight of evidence determination, and identification of reviews to which analyses used in the attainment demonstration have been subject.

Recommendations. States should address the 9 subject areas shown in Table 6.1 in the documentation accompanying an attainment demonstration. The documentation should contain a summary section which addresses issues shown in the table. More detailed information should be included in appendices, as necessary.

Table 6.1. Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone

Subject Area	Purpose of Documentation	Issues Included
Modeling/Analysis Protocol	Communicate scope of the analysis and document stakeholder involvement	Names of stakeholders participating in preparing and implementing the protocol; Types of analyses performed; Steps followed in each type of analyses; Days and domain considered.
Emissions Preparations and Results	Assurance of valid, consistent emissions data base. Appropriate procedures are used to derive emission estimates needed for air quality modeling.	Data base used and quality assurance methods applied; Data processing used to convert data base to model-compatible inputs; Deviations from existing guidance and underlying rationale; VOC, NOx, CO emissions by State/county for major source categories.
Air Quality/Meteorology Preparations and Results	Assurance that representative air quality and meteorological inputs are used in analyses	Extent of data base and procedures used to derive & quality assure inputs for analyses used in the weight of evidence determination; Departures from guidance and their underlying rationale. Performance of meteorological model if used to generate meteorological inputs to the air quality model.

Table 6.1. Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone (continued)

Subject Area	Purpose of Documentation	Issues Included
Performance Evaluation for Air Quality Model (and Other Analyses)	Show decision makers and the public how well the model (or other analyses) reproduced observations or otherwise performed on the days selected for analysis	Summary of observational data base available for comparison; Identification of performance tests used and their results; Ability to reproduce observed temporal and spatial patterns; Overall assessment of what the performance evaluation implies.
Diagnostic Tests	Ensure rationale used to adjust model inputs or to discount certain results is physically justified and the remaining results make sense	Results from application prior to adjustments; Consistency with scientific understanding and expectations; Tests performed, changes made and accompanying justification; Short summary of final predictions.

Table 6.1. Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone (continued)

Subject Area	Purpose of Documentation	Issues Included
Description of the Strategy Demonstrating Attainment	**Provide the EPA and the public an overview of the plan selected in the attainment demonstration.**	**Qualitative description of the attainment strategy;** **Reductions in VOC, NOx, and/or CO emissions from each major source category for each State/county from current (identify) emission levels;** **Clean Air Act mandated reductions and other reductions;** **Show predicted 8-hr site-specific future design values for the selected control scenario and identify any location which fails the screening test described in Section 3.4;** **Identification of authority for implementing emission reductions in the attainment strategy.** **Evidence that emissions remain at or below projected levels throughout the 3-year period used to determine future attainment.**
Data Access	**Enable the EPA or other interested parties to replicate model performance and attainment simulation results, as well as results obtained with other analyses.**	**Assurance that data files are archived and that provision has been made to maintain them;** **Technical procedures for accessing input and output files;** **Identify computer on which files were generated and can be read, as well as software necessary to process model outputs;** **Identification of contact person, means for downloading files and administrative procedures which need to be satisfied to access the files.**

Table 6.1. Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone (concluded)

Subject Area	Purpose of Documentation	Issues Included
Weight of Evidence Determination	Assure the EPA and the public that the strategy meets applicable attainment tests and is likely to produce attainment of the NAAQS within the required time.	Description of the modeled attainment test and observational data base used; Identification of air quality model used; Identification of other analyses performed; Outcome of each analysis, including the modeled attainment test; Assessment of the credibility associated with each type of analysis in this application; Narrative describing process used to conclude the overall weight of available evidence supports a hypothesis that the selected strategy is adequate to attain the NAAQS.
Review Procedures Used	Provide assurance to the EPA and the public that analyses performed in the attainment demonstration reflect sound practice	Scope of technical review performed by those implementing the protocol; Assurance that methods used for analysis were peer reviewed by outside experts; Conclusions reached in the reviews and the response thereto.

(This page is left blank intentionally)

7.0 References Cited In Part I And In Section 1.0

Blanchard, C.L. and P.M.Roth, (1997), *User's Guide Ozone M.A.P.P.E.R., Measurement-based Analysis of Preferences in Planned Emission Reductions, Version 1.1,* Final Report prepared for the U.S. EPA pursuant to Contract 68-D6-0064 Work Assignment, W.M.Cox, EPA Work Assignment Manager.

Blanchard, C.L., F.W.Lurmann, P.M.Roth, H.E.Jeffries and M.Korc, (1999), "The Use of Ambient Data to Corroborate Analyses of Ozone Control Strategies", *Atmospheric Environment, 33*, pp. 369-381.

Cardelino, C.A. and W.L.Chameides, (1995), "An Observation Based Model for Analyzing Ozone Precursor Relationships in the Urban Atmosphere", *J.Air & Waste Mgmt.Assoc 45*, pp.161-181.

Clinton, W.J., (July 16, 1997), Memorandum to the Administrator of the Environmental Protection Agency, Subject: "Implementation of Revised Air Quality Standards for Ozone and Particulate Matter".

Cox, W.M. and S.Chu, (1993), "Meteorologically Adjusted Ozone Trends in Urban Areas: A Probabilistic Approach", *Atmospheric Environment 27B*, (4), pp.425-434.

Cox, W.M. and S.Chu, (1996), "Assessment of Interannual Ozone Variation in Urban Areas from a Climatological Perspective", *Atmospheric Environment 30*, pp.2615-2625.

Deuel, H.P. and S.G.Douglas, (1998), *Episode Selection for the Integrated Analysis of Ozone, Visibility and Acid Deposition for the Southern Appalachian Mountains*, Draft Technical Report Systems Applications International, Inc. (SYSAPP-98/07), prepared for the Southern Appalachians Mountains Initiative.

Gao, D., W.R.Stockwell and J.B.Milford, (1996), "Global Uncertainty Analysis of a Regional Scale Gas-Phase Chemical Mechanism", *J.Geophys.Res. 101*, pp.9107-9119.

Henry, R.C., C.W.Lewis and J.F.Collins, (1994), "Vehicle-related Hydrocarbon Source Compositions from Ambient Data: The GRACE/SAFER Method", *Environmental Science and Technology 28*, pp.823-832.

Henry, R.C., (1997), "Receptor Model Applied to Patterns in Space (RMAPS) Part I. Model Description", *J.AWMA 47*, p.216.

Henry, R.C., (1997a), "Receptor Model Applied to Patterns in Space (RMAPS) Part II. Apportionment of Airborne Particulate Sulfur from Project MOHAVE", *J.AWMA 47*, p.220.

Henry, R.C., (1997b), "Receptor Modeling Applied to Patterns in Space (RMAPS) Part III. Apportionment of Airborne Particulate Sulfur in Western Washington State", *J.AWMA 47*, p.226.

Lefohn, A.S., D.S. Shadwick and S.D. Ziman, (1998), "The Difficult Challenge of Attaining EPA's New Ozone Standard", *Environmental Science and Technology, Policy Analysis, 32*, No.11, (June 1, 1998), pp.276A-282A.

Meyer, E.L., K.W.Baldridge, S.Chu and W.M.Cox, (1997), "Choice of Episodes to Model: Considering Effects of Control Strategies on Ranked Severity of Prospective Episode Days", *Paper 97-MP112.01*, Presented at 97th Annual AWMA Meeting, Toronto, Ontario, (June 1997).

Pacific Environmental Services, Inc., (1997), *Draft Technical Memorandum Analysis of 8-hour Ozone Values: 1980-1995*, Report prepared under EPA Contract No.68D30032, Work Assignment III-88, Edwin L. Meyer, Jr., Work assignment manager.

Reynolds, S.D., H.M.Michaels, P.M.Roth, T.W.Tesche, D.McNally, L.Gardner and G.Yarwood, (1997), "Alternative Base Cases in Photochemical Modeling: Their Construction, Role and Value", *Atmospheric Environment*, In press.

Sillman, S., (1997), "The Method of Photochemical Indicators as a Basis for Analyzing O3-NOx-ROG Sensitivity", NARSTO critical review paper, to be published in *Atmospheric Environment.*

Sillman, S., (1998), *Evaluating the Relation Between Ozone, NO_x and Hydrocarbons: The Method of Photochemical Indicators*, EPA/600R-98/022.

Sisler, J.F. (Editor), (1996), *Spatial and Seasonal Patterns and Long Term Variability of the Composition of the Haze in the United States: An Analysis of Data from the IMPROVE Network*, Cooperative Institute for Research in the Atmosphere Report ISSN: 0737-5352-32, Colorado State University, Fort Collins, CO.

U.S. EPA, (1993), *Volatile Organic Compound (VOC)/Particulate Matter (PM) Speciation Data System (SPECIATE), Version 1.5*, EPA/C-93-013.

U.S. EPA, (1994), *Guidance on Urban Airshed Model (UAM) Reporting Requirements for Attainment Demonstration*, EPA-454/R-93-056.

U.S. EPA, (1996), *Guidance on Use of Modeled Results to Demonstrate Attainment of the Ozone NAAQS*, EPA-454/B-95-007.

U.S. EPA, (1998a), *EPA Third-Generation Air Quality Modeling System, Models-3 Volume 9b Users Manual*, EPA-600/R-98/069(b).

U.S. EPA, (1998b), *Implementation Plan $PM_{2.5}$ Monitoring Program*, Available from EPA Internet website, ***"http://www.epa.gov/ttn/amtic/files/ambient/pm25/pmplan3.pdf"***.

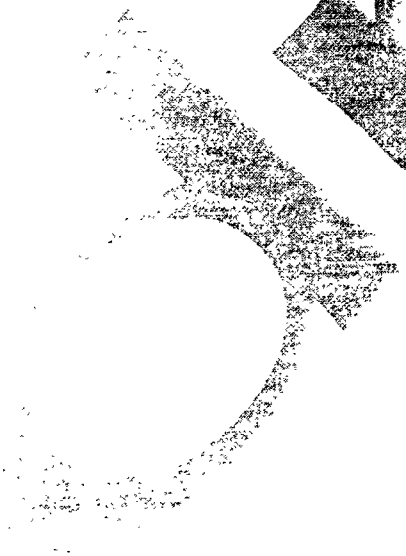
U.S. EPA, (1999a), *Implementation Guidance for the Revised Ozone and Particulate Matter (PM) National Ambient Air Quality Standards (NAAQS) and Regional Haze Program.*

Watson, J.G., (1997), "Observational Models: The Use of Chemical Mass Balance Methods", Prepared for the U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet website, ***"http://www.epa.gov/ttn/faca/stissu.html"*** and download file "source attribution workshop materials".

Yang, Y.J., W.R. Stockwell, J.B.Milford, (1995),"Uncertainties in Incremental Reactivities of Volatile Organic Compounds", *Environmental Science and Technology 29*, pp.1336-1345.

(This page is left blank intentionally)

Part II. How Should I Apply Air Quality Models To Produce Results Needed To Help Demonstrate Attainment?



(This page is left blank intentionally)

8.0 How Do I Apply Air Quality Models?-- An Overview

In Part I of this guidance, we described how to estimate whether a proposed control strategy will lead to attainment of the ozone NAAQS within a required time frame. We noted that air quality models play a major role in making this determination. We assumed that modeling had been completed, and discussed how to use the information produced. We now focus on how to apply models to generate the information used in the modeled attainment demonstration. The procedure we recommend consists of 8 steps:

1. formulate a conceptual description of an area's nonattainment problem;
2. develop a modeling/analysis protocol;
3. select an appropriate air quality model to use;
4. select appropriate meteorological episodes to model;
5. choose a modeling domain with an appropriate number of vertical layers and appropriately sized grid cells;
6. generate meteorological and air quality inputs to the air quality model;
7. generate emissions inputs to the air quality model;
8. evaluate performance of the air quality model and perform diagnostic tests.

In this Section (Section 8.0), we briefly describe each of these steps to better illustrate how they are interrelated. Because many of these steps require considerable effort to execute, States should take care to keep the appropriate U.S. EPA Regional Office(s) informed as they proceed. This will increase the likelihood of having an approvable attainment demonstration when the work is completed. Steps outlined in this Section are described in greater depth in Sections 9.0 - 16.0.

1. Formulate a conceptual description of an area's nonattainment problem. A State needs to have an understanding of the nature of an area's nonattainment problem before it can proceed with a modeled attainment demonstration. For example, it would be difficult to identify appropriate stakeholders and develop a modeling protocol without knowing whether resolution of the problem may require close coordination and cooperation with other nearby States.

The State containing the designated nonattainment area is expected to initially characterize the problem. This characterization provides a starting point for addressing steps needed to generate required information by those implementing the protocol. Several examples of issues addressed in the initial description of a problem follow. Is it a regional or local problem? Are factors outside of the nonattainment area likely to affect what needs to be done locally? Are monitoring sites observing violations located in areas where meteorology is complex or where there are large emission gradients? How has observed air quality responded to past efforts to reduce precursor emissions? Are there any ambient measurements suggesting which precursors and sources are important to further reduce ozone? What information might be needed from potential stakeholders? As many of the preceding questions imply, an initial conceptual description may be based largely on a review of ambient air quality data. Sometimes,

methods described in Section 4.0 (e.g., trend analysis, observational models) may be used. Other times, these types of analyses may be deferred until after a team is in place to develop and implement steps following a modeling/analysis protocol. The initial conceptual picture may be based on less resource intensive analyses of available data.

2. Develop a modeling/analysis protocol. A protocol describes how modeling will be performed to support a particular attainment demonstration. Its direction and participating stakeholders are influenced by the previously developed conceptual description of the problem to be resolved. The protocol outlines methods and procedures which will be used to perform the subsequent 6 steps needed to generate the modeling results and to subsequently apply the modeled attainment and screening tests as well as other corroborating analyses in a weight of evidence determination. It does this by: a) identifying those responsible for implementing the modeling, b) identifying those who will review each step as it occurs, c) identifying procedures to be used to consider input/suggestions from those potentially affected by the outcome (i.e., "stakeholders"), and d) outlining how decisions will be made concerning technical analyses needed to complete each step in the modeling procedure. In short, the protocol defines the "game plan" and the "rules of the game".

3. Select an appropriate model for use. This step includes reviewing air quality data to gain insight about the nature of a nonattainment area's ozone problem, reviewing rules established in the *Guideline for Air Quality Models* (U.S. EPA, 1999b), and considering experience/expertise of those performing the modeling. Identifying the air quality model to be used is an early step in the process, since it may affect how emissions and meteorological information are input to the model. It could also affect size of the area modeled and choice of the horizontal/vertical resolution considered.

4. Select appropriate meteorological episodes to model. Like the preceding step, this step requires review of available air quality data. It also requires a thorough understanding of the form of the national ambient air quality standard and of the modeled attainment test described in Sections 3.1 and 3.2. Finally, it requires a review of meteorological conditions which have been observed to accompany monitored exceedances of the concentration specified in the NAAQS (i.e., $\geq$ 85 ppb). The object of these reviews is to select episodes which a) include days with observed concentrations close to site-specific design values so that all sites with current design values $\geq$ 75 ppb can be considered in the modeled attainment test, and b) reflect a variety of meteorological conditions which have been commonly observed to accompany monitored exceedances. This latter objective is desirable, because it adds confidence that a proposed strategy will work under a variety of conditions.

5. Choose a modeling domain with an appropriate number of vertical layers and appropriately sized grid cells. Appropriate domain size is influenced by the choice of episodes modeled. Meteorological and air quality (i.e., ozone) data corresponding to these episodes and, if applicable, to other, plausible episodes, need to be reviewed prior to choosing size of the area modeled. Presence of topographical features or mesoscale meteorological features (e.g., land/sea

breeze) near or in the nonattainment area of principal interest are factors to consider in choosing size of individual grid cells and the number of required vertical layers for that portion of the modeling grid. Another factor affecting choice of grid cell size is the available spatial detail in the emissions data used as input to an emissions model. Finally, feasibility of managing large data bases and resources needed to estimate meteorological inputs and air quality in many grid cells are factors which cannot be ignored in choosing size of a domain and its grid cells.

6. Generate meteorological and air quality inputs to the air quality simulation model. Unlike emissions, meteorological inputs remain constant during "base case current" and "future" periods simulated with the air quality model. Nevertheless care needs to be taken in specifying these, as they may affect relationships predicted between ozone and emissions. Ozone modeling may have to consider large geographical areas in many instances. Further, past modeling has shown that meteorological conditions aloft can have an important effect on predicted ozone. Finally, meteorological monitoring is relatively sparse outside of cities and, especially, aloft. Thus, we recommend that meteorological models ordinarily be used to generate meteorological inputs. Application of meteorological models and choice of model grid resolution in the preceding step are closely related. Meteorological conditions near the area which is the focus of an attainment demonstration may dictate the required spatial resolution. On the other hand, cost and data management difficulties increase greatly for finely resolved grids. Thus, those implementing the protocol will likely be faced with a tradeoff between cost/feasibility of running air quality and meteorological models and resolution at which it might be most desirable to treat dispersion of nearby emissions.

Air quality inputs consist of initial conditions and boundary conditions to the model domain. Importance of initial conditions should be diminished by beginning a simulation at a time prior to the period which is of interest. Nature of boundary conditions is an important factor in deciding how large to make the size of the area modeled. The most satisfactory way to generate future boundary conditions is through use of a regional air quality model. Therefore, those implementing the protocol will once again be faced with a tradeoff between cost/feasibility of data base management vs. a desire to limit the importance of an arbitrarily specified input to the modeling exercise.

7. Generate emissions inputs to the air quality simulation model. Emissions are the central focus in a modeled attainment demonstration. That is, they are the only input to an air quality model which those implementing the protocol can control. Hence, they are the major input which gets changed between the present and future. Emissions which are input to an air quality model are generated using an emissions model. Applying such a model is as complicated as the air quality model itself, and demands at least as much attention. In current emissions models, emissions from some of the major source categories of ozone precursors are affected by meteorological conditions. This requires an interface between meteorological inputs and emissions. Emissions which are input to the air quality model are also affected by the latter's horizontal/vertical resolution and, of course, the size of the area modeled. In short, treatment of emissions is a central and complex one which, itself, involves several steps. These include

deriving emission inventories, quality assuring results, applying results in an emission model(s), and (again) quality assuring results. Emission inputs may be needed for as many as 3 periods: (1) a "base case period" corresponding to that of the selected episodes, (2) a "current period", corresponding to that represented by the current monitored design value, and (3) a future period, corresponding to a time two years prior to the required attainment date.

8. Evaluate performance of the air quality simulation model and perform diagnostic tests. To an important extent, credibility of a modeled attainment test's results and other modeled outputs is affected by how well the model replicates observed air quality. Evaluating model performance and conducting diagnostic tests depend on prior definition of the modeling exercise and specification of model inputs. Hence, this is generally the last step prior to using the model to support an attainment demonstration, as described in Part I.

In the past, performance evaluation has relied almost exclusively on numerical tests comparing predicted and observed ozone, or visual inspection of predictions and observations. These are still important tools. However, photochemical grid models have many inputs, and it is possible to get similar predicted ozone concentrations with different combinations of these inputs. There is no guarantee that ozone will respond the same way to controls with these different combinations of inputs. Thus, we place greater emphasis on additional kinds of tests than was true in past guidance. These include use of precursor observations, use of indicator species, use of corroborative analyses with observational models and use of retrospective analyses.

Diagnostic tests are separate simulations which are performed to determine the sensitivity of a model's ozone predictions to various inputs to the model. This can be done for a variety of purposes, including selection of effective control strategies, prioritizing inputs needing greatest quality assurance and assessing uncertainty associated with model predictions. In performing such tests, States should remember how model results are used in the modeled attainment test recommended in Section 3.0. In general, model results are used in a relative rather than absolute sense. In particular, the modeled attainment test requires use of relative reduction factors (RRF), generated by models. Thus, diagnostic tests should be used to consider how RRF, as well as absolute ozone predictions, are affected by changes to model inputs.

Recommendations. States should follow eight steps in applying models to generate information required for use in modeled attainment demonstrations.

1. **Formulate a conceptual description of an area's nonattainment problem.**
2. **Develop a modeling/analysis protocol.**
3. **Choose an appropriate model.**
4. **Choose appropriate episodes.**
5. **Choose a modeling domain with an appropriate number of vertical layers and appropriately sized grid cells.**
6. **Generate appropriate meteorological and air quality inputs.**

7. Generate quality assured emissions inputs.
8. Evaluate model performance and undertake diagnostic tests.

Execution of subsequent steps should be performed in accordance with procedures identified in the protocol. Rationale and outcome of the steps should be documented as described in Section 6.0. To increase the likelihood of an approvable demonstration, States should carefully coordinate development and execution of steps with the appropriate U.S. EPA Regional Office(s).

(This page is left blank intentionally)

9.0 How Do I Get Started?

A State should start developing information to support a modeled attainment demonstration by assembling and reviewing available air quality, emissions and meteorological data. Current design values should be calculated at each ozone monitoring site, as described in Section 3.1. If past modeling has been performed, the emission scenarios examined and air quality predictions may also be useful. Readily available information should be used by a State to develop an initial conceptual description of the nonattainment problem in the area which is the focus of a modeled attainment demonstration. A conceptual description is instrumental for identifying potential stakeholders and for developing a modeling/analysis protocol. It may also influence a State's choice of air quality model, modeling domain, grid cell size, priorities for quality assuring and refining emissions estimates and choice of initial diagnostic tests to identify potentially effective control strategies. In general, a conceptual description is useful for helping a State to identify priorities and allocate resources in performing a modeled attainment demonstration.

In this Section, we identify key parts of a conceptual description. We then present examples of analyses which could be used to describe each of these parts. We note that initial analyses may be complemented later by additional efforts performed by those implementing the protocol, and that many of the analyses we describe would be more convincing with improved data bases.

9.1 What Is A "Conceptual Description"?

A "conceptual description" is a qualitative way of characterizing the nature of an area's nonattainment problem. It is best described by identifying key components of a description. Examples are listed below. The examples are not necessarily comprehensive. There could be other features of an area's problem which are important in particular cases. For purposes of illustration later in the discussion, we have answered each of the questions posed below. Our responses appear in parentheses.

-1. Is the nonattainment problem primarily a local one, or are regional factors important?

(Surface measurements suggest transport of ozone close to 84 ppb is likely. There are some other nonattainment areas not too far distant)

-2. Are ozone and/or precursor concentrations aloft also high?

(There are no such measurements.)

-3. Do violations of the NAAQS occur at several monitoring sites throughout the nonattainment area, or are they confined to one or a small number of sites in proximity to one another?

(Violations occur at a limited number of sites, located throughout the area.)

-4. Do occasions in which observed 8-hour daily maximum ozone concentrations exceed 84 ppb occur often or just on a few occasions?

(This differs for different monitors from 4 times up to 12 times per year.)

-5. When 8-hour daily maxima in excess of 84 ppb occur, is there an accompanying characteristic spatial pattern of these, or is there a variety of spatial patterns?

(A variety of patterns is seen.)

-6. Do monitored violations occur at locations subject to mesoscale wind patterns (e.g., at a coastline) which may differ from the general wind flow?

(No.)

-7. Have there been any recent major changes in emissions of VOC or NOx in or near the nonattainment area? What?

(Yes, 4 measures believed to result in major reductions in VOC have been implemented in the last 5 years.)

-8. Are there discernible trends in design values or other air quality indicators which have accompanied a change in emissions?

(Yes, design values have decreased by about 10% at 4 sites, smaller or no reduction is seen at 3 other sites.)

-9. Is there any apparent spatial pattern to the trends in design values?

(No.)

-10. Have ambient precursor concentrations or measured VOC species profiles changed?

(There are no measurements.)

-11. What past modeling has been performed and what do the results suggest?

(A regional modeling analysis has been performed. Two emission scenarios were modeled: current emissions and a substantial reduction in NOx emissions throughout the regional domain. Reduced NOx emissions led to substantial predicted reductions in 8-hour daily maximum ozone in most locations, but changes

near the most built-up area in the nonattainment area in question were small or nonexistent.)

-12. Are there any distinctive meteorological measurements at the surface or aloft which appear to coincide with occasions with 8-hour daily maxima greater than 84 ppb?

(Other than routine soundings taken twice per day, there are no measurements aloft. There is no obvious correspondence with meteorological measurements other than daily maximum temperatures are always ≥ 85F on these days.)

Using responses to the preceding questions in this example, it is possible to construct an initial conceptual description of the nonattainment area's ozone problem. First, responses to questions 1 and 11 suggest there is a significant regional component to the area's nonattainment problem. Second, responses to questions 3, 4, 7 and 8 indicate there is an important local component to the area's nonattainment problem. The responses to questions 4, 5 and 12 indicate that high ozone concentrations may be observed under several sets of meteorological conditions. The responses to questions 7, 8 and 11 suggest that ozone in and near the nonattainment area may be responsive to both VOC and NOx controls and that the extent of this response may vary spatially. The response to question 6 suggests that it may be appropriate to develop a strategy using a model with 12 km grid cells rather than finer resolution and later check its adequacy for finer resolution on a limited basis.

The preceding conceptual description implies that the State containing the nonattainment area in this example will need to involve stakeholders from other, nearby States to develop and implement a modeling/analysis protocol. It also suggests that a nested regional modeling analysis will be needed to address the problem. Further, it may be necessary to model several distinctive types of episodes and additional analyses will be needed to select episodes. Finally, sensitivity (i.e., diagnostic) tests will be needed to assess effects of reducing VOC and NOx emissions separately and in tandem.

It should be clear from the preceding example that the initial conceptual description of an area's nonattainment problem may draw on readily available information and need not be detailed. It is intended to help launch development and implementation of a modeling/analysis protocol in a productive direction. It will likely be supplemented by subsequent, more extensive modeling and ambient analyses performed by or for those implementing the modeling/analysis protocol discussed in Section 10.0.

Recommendations. States should begin an analysis to support a modeled attainment demonstration by developing a conceptual description of an area's nonattainment problem. This description is based on use of readily available air quality, meteorological and emissions information. It may be refined later as additional analyses are performed by those implementing the modeling/analysis protocol.

9.2 What Sorts Of Analyses Might Be Useful For Developing And Refining A Conceptual Description?

Questions like those posed in Section 9.1 can be addressed using a variety of analyses ranging in complexity from an inspection of air quality data to sophisticated mathematical analyses. We anticipate the simpler analyses will often be used to develop the initial conceptual description. These will be followed by more complex approaches or by approaches requiring more extensive data bases as the need later becomes apparent. In the following paragraphs, we revisit key parts of the conceptual description identified in Section 9.1. We note analyses which may help to develop a description of each part. The list serves as an illustration. It is not necessarily exhaustive.

1. Is regional transport an important factor affecting the nonattainment area?

-Are there other nonattainment areas within a day's transport of the nonattainment area?

-Do "upwind" 8-hour daily maximum ozone concentrations approach or exceed 84 ppb on some or all of the days with observed 8-hour daily maxima > 84 ppb in the nonattainment area?

-Are there major sources of emissions upwind?

-What is the size of the downwind/upwind gradient in 8-hour daily maximum ozone concentrations compared to the upwind values?

-Do ozone concentrations aloft but within the planetary boundary layer approach or exceed 84 ppb at night or in the morning hours prior to breakup of the nocturnal surface inversion?

-Is there a significant positive correlation between observed 8-hour daily maximum ozone concentrations at most monitoring sites within or near the nonattainment area?

-Is timing of high observed ozone consistent with impacts estimated from upwind areas using trajectory models?

-Do available regional modeling simulations suggest that 8-hour daily maximum ozone concentrations within the nonattainment area respond to regional control measures?

2. What or how many meteorological episodes lead to high ozone?

-Examine spatial patterns of 8-hour daily maxima occurring on each day for which a value > 84 ppb occurs to try to identify a limited number of distinctive patterns.

-Review synoptic weather charts for days having observed concentrations > 84 ppb to identify classes of synoptic scale features corresponding to high observed ozone.

-Perform statistical analyses between ozone or 8-hour daily maximum ozone and meteorological measurements at the surface and aloft to identify distinctive classes of days corresponding with observed daily maxima > 84 ppb.

-Apply indicator species methods such as those described by Sillman (1998) and Blanchard, et al. (1999) at sites with appropriate measurements on days with ozone exceedances. Identify classes of days where further ozone formation appears limited by available NOx vs. classes of days where further ozone formation appears limited by available VOC.

3. Is ozone limited by availability of VOC, NOx or combinations of the two? What sorts of source categories may be important?

-What are the major source categories of VOC and NOx and what is their relative importance in the most recent inventory?

-Review results from past modeling analyses to assess the likelihood that ozone in the nonattainment area will be more responsive to VOC or NOx controls. Do conclusions vary for different locations?

-Apply indicator species methods at sites with appropriate measurements to assess whether maximum observed ozone at these sites is limited by the availability of VOC or NOx. Do conclusions differ for different days?

-Apply receptor modeling approaches such as those described by Watson (1997), Henry, et al. (1994) and Henry (1997, 1997a, 1997b) to identify source categories contributing to ambient VOC on days with high observed ozone. Do conclusions differ on days when measured ozone is not high?

In the preceding example, we listed analyses to help describe three major components of a conceptual description so that the easiest analysis, relying on available data, appears first. The more analyses a State is able to perform, the more complete and more accurate the description of an area's nonattainment problem may be. As noted in Section 5.0, the most complete description will depend on use of refined data bases which supplement routinely collected data. For example, statistical models between meteorological variables and observed ozone will probably better describe relationships if meteorological measurements are available from aloft.

Some of the analyses may be identified as desirable as issues arise in implementing a modeling/analysis protocol. Their function is to channel resources available to support modeled attainment demonstrations onto the most productive paths possible. They also provide other pieces of information which can be used to reinforce conclusions reached with an air quality model, or cause a reassessment of assumptions made previously in applying the model. As noted in Section 4.0, corroboratory analyses may also be used in a weight of evidence determination to help assess whether a simulated control strategy is sufficient to meet the NAAQS.

Recommendations. States should analyze ambient air quality, meteorological and emissions data in concert with an air quality modeling analysis. These analyses perform at least 3 functions. First, they are needed to help develop a conceptual description of a nonattainment area's problem. Second, they help guide application of a model in an air quality modeling analysis. Third, analysis of air quality, meteorological and emissions data generates corroborative information which may confirm conclusions drawn with an air quality model or cause some of the underlying assumptions in the modeling to be reexamined.

10.0 What Does A Modeling/Analysis Protocol Do, And What Does Developing One Entail?

Developing and implementing a modeling/analysis protocol is a very important part of an acceptable modeled attainment demonstration. Much of the information in U.S. EPA (1991) regarding modeling protocols remains applicable. States should review the 1991 guidance on protocols. In this document, we have revised the name of the protocol to "Modeling/Analysis Protocol" to emphasize that the protocol needs to address all types of analyses considered in a weight of evidence determination, not just modeling.

10.1 What Is The Protocol's Function?

The most important function of a protocol is to serve as a means for planning and communicating how a modeled attainment demonstration will be performed *before* it occurs. The protocol is the means by which States and other stakeholders can assess applicability of default recommendations described herein and develop alternatives. A good protocol should lead to widespread participation in developing the demonstration. It should also reduce risk of spending time and resources on efforts which the appropriate U.S. EPA Regional Office(s) believes are unproductive or inconsistent with Agency policy.

The protocol also serves several important, more specific functions. First, it identifies who will be helping the State or local air quality agency (generally the lead agency) to undertake or evaluate analyses needed to support a defensible demonstration (i.e., the stakeholders). Second, it identifies how communication will occur among stakeholders to develop consensus on various issues. Third, it identifies methods and procedures used to support the demonstration. Fourth, the protocol describes the review process applied to key steps in the demonstration. Fifth, it describes how changes in methods and procedures or in the protocol itself are agreed upon and communicated with stakeholders and the appropriate U.S. EPA Regional Office(s). Major steps taken in implementing the protocol should be discussed with the appropriate U.S. EPA Regional Office(s) as they are being decided. States should update the protocol as major decisions are made concerning forthcoming analyses.

10.2 What Subjects Should Be Addressed In The Protocol?

States should address the following subjects in their modeling/analysis protocol:

1. Stakeholders participating in the process.

2. Management/communication procedures used, including those to amend the protocol.

3. Choice of the air quality simulation model to be used and how it meets requirements in 40CFR51, Appendix W for using "alternative" models.

4. Assurance that proposed modeling procedures have been scientifically peer reviewed and plans

for technical review of how procedures are used in the specific application and the resulting outputs.

5. Types of analyses included in the weight of evidence determination, if used.

6. Outcomes for each analysis which will be considered consistent with suggesting a selected strategy will meet the NAAQS.

7. Data base used to support air quality modeling and other types analyses used in a weight of evidence determination.

8. Rationale for choice of air quality and emissions model and choice of method for generating meteorological inputs

9. Methods used to quality assure emissions inputs

10. Domain size and spatial resolution to be used.

11. Criteria/goals in selecting periods to model and process to be used in selecting episodes.

12. Performance evaluation procedures and additional diagnostic tests planned.

13. Outcomes in the modeled attainment and screening tests as well as results of analyses to be used in a broader weight of evidence determination.

14. Procedures to be used to archive, document and report results.

15. Identification of specific deliverables and schedule for delivery to the appropriate U.S. EPA Regional Office.

Recommendations. States should prepare a modeling/analysis protocol as part of an acceptable demonstration of attainment. Generally, procedures recommended in the 1991 guidance and followed for the 1994 ozone SIP revisions are appropriate. These procedures should be augmented to include a discussion of all analyses to be included in the weight of evidence determination, not just modeling. The protocol should also include provision for review of key parts of the analysis and data base underlying the attainment demonstration. The protocol should be kept up to date to reflect major changes in initial plans.

11.0 What Should I Consider In Choosing An Air Quality Model?

Photochemical grid models are, in reality, *modeling systems* in which an emissions model, a meteorological model and an air chemistry/deposition model are applied. In this guidance, we use the term "air quality model" to mean a gridded photochemical modeling system. Some modeling systems are modular, at least in theory. This means that it is possible to substitute alternative emissions or meteorological models within the modeling system. Often however, choice of an emissions or meteorological model or their features is heavily influenced by the chosen air quality model (i.e., an effort is needed to develop software to interface combinations of components differing from the modeling system's default combination). Thus, choice of an appropriate air quality model is among the earliest decisions to be made by those implementing the protocol. In this section, we identify a set of general requirements which an air quality model should meet to qualify for use in an attainment demonstration for the 8-hour ozone NAAQS. We then identify several factors which will help in choosing among qualifying air quality models for a specific application. We conclude this Section by identifying several air quality models which are available for use in attainment demonstrations. Meteorological and emissions models are discussed in Sections 14.0 and 15.0, respectively.

11.1 What Prerequisites Should An Air Quality Model Meet To Qualify For Use In An Attainment Demonstration?

A model should meet several general criteria for it to be a candidate for consideration in an attainment demonstration. These general criteria are consistent with requirements in 40CFR Part 51, Appendix W (i.e., the "*Model Guideline*") to be proposed in 1999. Note that, unlike in previous guidance (U.S. EPA, 1991), we are not recommending a specific model for use in the attainment demonstration for the 8-hour NAAQS for ozone. At present, there is no single model which has been extensively tested and shown to be clearly superior or easier to use than several alternatives. Thus, at this time, we do not anticipate that the next revision to 40CFR Part 51 Appendix W will identify a "preferred model" for use in attainment demonstrations of the 8-hour NAAQS for ozone. Using language in 40CFR Part 51 Appendix W, States should consider nested regional air quality models or urban scale air quality models as "alternative models" for ozone.

The U.S. EPA has invested considerable effort to develop a nested regional model (CMAQ) within a modeling system called "MODELS3" (U.S. EPA, 1998a). The U.S. EPA will provide support, in the form of documentation, user's guides, computer codes, updates, training and limited troubleshooting for the CMAQ model. The CMAQ model is designed to address ozone, $PM_{2.5}$ and regional haze-related applications. However, this model has not, as yet, been shown to be clearly superior or easier to use than available alternatives. Thus, use of the CMAQ model is subject to the same review criteria as other "alternative models" proposed to support an attainment demonstration of the 8-hour ozone NAAQS.

"Alternative models" may be used if they are non-proprietary. A "non-proprietary"

model is one whose source code is available for free or for a reasonable cost. Further, the user must be free to revise the code to perform diagnostic analyses and/or to improve the model's ability to describe observations in a credible manner. Several additional prerequisites should be met for an "alternative model" to be used to support a modeled attainment demonstration.

(1) It should have received a scientific peer review.

(2) It should be applicable to the specific application on a theoretical basis.

(3) It should be used with a data base which is adequate to support the application.

(4) It should have performed in past applications in such a way that estimates are not likely to be biased low.

(5) It should be applied consistently with a protocol on methods and procedures.

An air quality model may be considered to have undergone a "scientific peer review" if each of the major components of the modeling system (i.e., air chemistry/deposition, meteorological and emissions models) has been described and tested, and the results have been documented and reviewed by one or more disinterested third parties. We believe that it should be the responsibility of the model developer or group which is applying an air quality model on behalf of a State to document that a "scientific peer review" has occurred. States should then reference this documentation to gain acceptance of an air quality model for use in a modeled attainment demonstration.

Should the U.S. EPA identify a "preferred model" at some future date, an "alternative model" may still be used in a subsequent application if it is shown to be more appropriate for the specific application. This could be demonstrated by side by side comparisons of predictions obtained with the "preferred" and "alternative" models with observations. While such comparisons may be desirable, they are not necessarily required. Criteria described in Section 11.2 may be used to show that an "alternative model" is more appropriate than a "preferred model" for a specific application.

Recommendations. For an air quality model to qualify as a candidate for use in an attainment demonstration of the 8-hour ozone NAAQS, a State needs to show that it meets several general criteria.

1. The model has received a scientific peer review.

2. The model can be demonstrated applicable to the problem on a theoretical basis.

3. Data bases needed to perform the analysis are available and adequate.

4. Available past appropriate performance evaluations have shown the model is not biased toward underestimates.

5. A protocol on methods and procedures to be followed has been established.

6. The developer of the model must be willing to make the source code available to users for free or for a reasonable cost, and the model cannot otherwise be proprietary.

11.2 What Factors Affect My Choice of A Model For A Specific Application?

States should consider several factors as criteria for choosing a qualifying air quality model to support an attainment demonstration for the 8-hour ozone NAAQS. These factors are: (1) nature of the observed air quality problem; (2) documentation and past track record of candidate models in similar applications; (3) experience of staff and available contractors; (4) required time and resources vs. available time and resources; (5) in the case of regional applications, consistency with regional models applied in adjacent regions. The first of these factors is used to identify attributes needed for a model to be chosen. Factors (2)-(5) are used to help choose among candidate models having these attributes. Finally, before results of a selected model can be used in an attainment demonstration, the model should be shown to perform satisfactorily using the data base available for the specific application.

Nature of the observed air quality problem. This is the most important criterion for selecting an appropriate model. Prior to selecting a model to use in an attainment demonstration, we recommend that those implementing the protocol review available air quality, meteorological and emissions data, and take account of the geographic location of the nonattainment area(s) relative to that of precursor emissions. Section 9.0 identifies some types of analyses which may be useful for developing a conceptual description of an area's nonattainment problem.

States should undertake this review to decide whether it is best to use an urban scale photochemical grid model (e.g., domain size ~ 200-300 km on a side) or a regional photochemical grid model (e.g., domain size ~1000 km or more on a side) with or without nesting. Choice between an urban scale and regional application depends on answers to several questions

1. Is transport of ozone (or precursors) into the nonattainment area a major contributor to an area's ozone problem?

2. Are nonattainment areas sufficiently numerous and in relatively close proximity so that it is more efficient to estimate control requirements for several nonattainment areas simultaneously?

3. Is the nonattainment area located near major sources of anthropogenic precursors and/or topographical features requiring fine scale resolution to adequately characterize wind flow?

Answers to the preceding questions require a case by case analysis of available air quality, emissions and meteorological data. Generally however, we anticipate that an urban scale model may suffice for "isolated" nonattainment areas (e.g., in the West, outside of California). Locations subject to transported ozone well above natural background (i.e., 8-hr. daily maximum "natural" background is ~ 40-50 ppb) may need to use a regional model. If there are major concentrations of anthropogenic precursor emissions within about 75 km of the area of concern, an urban scale or nested regional model (incorporating a more finely resolved grid over a limited area) is advisable. An urban scale or nested regional model is also recommended if receptor sites of interest are located near a major body of water.

Documentation and Past Track Record of Candidate Models. For a model to be used in an attainment demonstration, evidence should be presented that it has been found acceptable for estimating hourly or 8-hourly ozone concentrations. Preference should be given to models exhibiting satisfactory past performance under a variety of conditions. Finally, a user's guide (including a benchmark example and outputs) and technical description of the model should be available.

Experience of Staff and Available Contractors. This is a legitimate criterion for choosing among several otherwise acceptable alternatives. The past experience might be with the air quality model itself, or with a meteorological or emissions model which can be more readily linked with one candidate air quality model than another.

Required vs. Available Time and Resources. This is a legitimate criterion provided the first two criteria are met.

Consistency of a Proposed Model with Models Used in Adjacent Regions. This criterion is applicable for regional model applications. If candidate models meet the other criteria, this criterion should be considered in choosing a model for use in a regional or nested regional modeling application.

Demonstration that an "Alternative Model" is Appropriate for the Specific Application. If an air quality model meets the prerequisites identified in Section 11.1, a State may use the factors described in this section (Section 11.2) to show that it is appropriate for use in a specific application. Choice of an "alternative model" for use in a specific attainment demonstration of the 8-hour NAAQS for ozone needs to be reviewed by the appropriate U.S. EPA Regional Office and by the U.S. EPA Model Clearinghouse.

Satisfactory Model Performance in the Specific Application. Prior to use of a selected model's results in an attainment demonstration, it should be shown to perform adequately in the specific application. Means for evaluating model performance are discussed in Section 16.0.

Recommendations. States should first determine what attributes are needed for a qualifying model to address a nonattainment area's ozone problem, and then choose among models possessing these attributes. Five factors should be considered in selecting an air quality model for a specific application. Selection of an air quality model should be concurred with by the appropriate U.S. EPA Regional Office and U.S. EPA Model Clearinghouse. The five factors are listed approximately in order of importance.

1. Nature of the air quality problem leading to nonattainment of the ozone NAAQS should first be assessed, and the selected model should have attributes and capabilities consistent with the perceived nature of the problem.

2. Availability, documentation and past performance should be satisfactory.

3. Relevant experience of available staff and contractors should be consistent with choice of a model.

4. Time and resource constraints may be considered.

5. Consistency of the model with what was used in adjacent regional applications should be considered.

Prior to using model results in a specific attainment demonstration, a State should show that the model performs adequately in replicating base case observations available for that demonstration.

11.3 What Are Some Examples Of Air Quality Models Which May Be Considered?

Table 11.1 lists several current generation air quality models which have been used to simulate ambient ozone concentrations. The list is not intended to be comprehensive. Exclusion of a model from the list does not necessarily imply that it cannot be used to support a modeled attainment demonstration for the ozone NAAQS. By the same token, inclusion on the list does not necessarily imply that a model may be used for a particular application. States should follow the guidance in Sections 11.1 and 11.2 in selecting an air quality model for a specific application.

Table 11.1. Some Air Quality Models Used To Model Ozone

Air Quality Model	References	Sponsors of Past Applications
CALGRID	Scire, et al. (1989)	Massachusetts Division of Air Quality Control (New England)
CAMx	Environ (1997)	Texas Natural Resources Conservation Commission (SE Texas and environs), Kansas Dept. Of Health & Environment (Kansas City and environs)
CMAQ	U.S. EPA (1998a)	U.S. EPA Office of Research & Development (eastern U.S. with fine grids in northeast corridor and at Nashville and environs)
MAQSIP	MCNC (1999) Odman, et al. (1996)	North Carolina Division of Environmental Management (Charlotte, most of NC and parts of surrounding States)
SAQM	Chang, et al., (1997) CARB (1996)	California Air Resources Board (San Joaquin Valley)
UAMV	Systems Applications International (1996)	LADCO (eastern U.S. with focus on States bordering Lake Michigan), New York Dept. Of Environmental Conservation (eastern U.S. with focus on northeast corridor)
URM	Kumar, et al., (1996)	Georgia Inst. Of Technology, Dr. A.G.Russell (northeastern U.S., Southern Appalachian Mountain Initiative, southern California)

12.0 How Do I Decide Which Meteorological Episodes To Model?

At a minimum, four criteria should be used to select episodes which are appropriate to model. First, choose a mix of episodes reflecting a variety of meteorological conditions which frequently correspond with observed 8-hour daily maxima > 84 ppb at different monitoring sites. Second, model periods in which observed 8-hour daily maximum concentrations are close to average 4th high 8-hour daily maximum ozone concentrations. Third, model periods for which extensive air quality/meteorological data bases exist. Fourth, model a sufficient number of days so that the modeled attainment test applied at each monitor violating the NAAQS is based on several days. The four criteria may often conflict with one another. For example, there may only be a limited number of days with intensive data bases, and these may not cover all of the meteorological conditions which correspond with monitored ozone concentrations close to site-specific design values during the base period. Thus, tradeoffs among the four primary criteria may be necessary in specific applications.

Those implementing the modeling/analysis protocol may use secondary episode selection criteria on a case by case basis. For example, prior experience modeling an episode, may result in its being chosen over an alternative. Another consideration should be to choose episodes occurring during the 3-year period which serves as the basis for the current monitored design value. As we note in Section 3.3, this could save some modeling resources/effort. A third consideration should be to try to ensure that episodes are chosen so that there are several days with monitored ozone concentrations near the site-specific design value at *each* monitoring site in a nonattainment area. If observed 8-hour daily maxima > 84 ppb occur on weekends, weekend days should be included within some of the selected episodes. If a State chooses to model several nonattainment areas simultaneously (e.g., with a nested regional model), a fifth secondary criterion is to choose episodes containing days of common interest to different nonattainment areas.

In this Section, we first discuss each of the four identified primary criteria for choosing meteorological episodes to model. We then discuss secondary criteria, which may be important in specific applications.

12.1 What Are The Most Important Criteria For Choosing Episodes?

Choose a mix of episodes which represents a variety of meteorological conditions which frequently correspond with observed 8-hour daily maxima exceeding 84 ppb. This criterion is important, because we want to be assured that a control strategy will be effective under a variety of conditions leading to ozone concentrations near current site-specific design values at sites where the NAAQS is violated. We believe the most important indicator of variety is differing wind fields. This affects source/source and source/receptor orientations and, therefore, the effectiveness of a strategy.

Those implementing the modeling/analysis protocol should describe the rationale for

distinguishing among episodes which are modeled. The selection may reflect a number of area specific considerations. Qualitative procedures such as reviewing surface and aloft weather maps, observed or modeled wind patterns may suffice for distinguishing episodes with distinctively different meteorological conditions. More quantitative procedures, such as a CART analysis, to identify distinctive groupings of meteorological/air quality parameters corresponding with high 8-hour daily maxima for ozone, may sometimes be desirable. An example of a CART analysis applied to select episodes is described by Deuel, et al. (1998).

Choose episodes having some days with monitored 8-hour daily maxima close to observed average 4th high daily maximum ozone concentrations. In Figure 3.3, we saw that, at any given site, the relative reduction factor (RRF) used in the test is not independent of predicted current 8-hour daily maxima when these are below about 70 ppb. However, Figure 3.3 reflects relationships between current predicted 8-hour daily maxima and RRF when future/current modeled concentration ratios are averaged over 10 days. It may not be practical to simulate enough days so that the test applied at each site reflects mean responses from as many as 10 days. Thus, we want to use episodes whose severity is comparable to that implied by the form of the NAAQS (i.e., an episode whose severity is exceeded, on average, about 3 times/year at the time of the selected episode). Note that we said, "at the time of the selected episode" (i.e., the "base case period") rather than "current period" in the preceding sentence. The objective is to choose episodes with days which are approximately as severe as the average 4th high 8-hour daily maximum concentration specified in the NAAQS.

Air quality measurements recorded during the base case period can be used to characterize episode severity. This is done by selecting a 3-year period which "straddles" a modeled episode. For example, if an episode from 1995 were modeled, we recommend looking at measured 8-hour daily maxima at each site in the nonattainment area during 1994-1996. Using this information it should be possible to assess the relative severity of the days chosen for modeling at each site. Limiting this characterization to the three years straddling an episode avoids problems posed by long term trends in emissions in assessing episode severity. However, it leaves unanswered the question of whether the 3-year period selected to assess severity of a modeled day is typical or atypical. If there is an underlying long term trend in ambient ozone attributable to meteorological cycles or other causes, it may not be appropriate to compare different 3-year periods with one another using air quality observations. Thus, if one uses a 10-year old episode with an exceptional data base, there is greater uncertainty in ranking its severity relative to the current period of interest than if the episode were drawn from the current period.

The problem of dealing with longer term variations in meteorological conditions producing high ozone can be reduced by assessing the potential of meteorological conditions to form high ozone in concert with a climatological data base. An example of such an approach is described in Cox, et al., (1996). If such an analysis shows that the 3-year periods straddling each selected episode day and the most recent 3-year period are not an extreme ones, this supports using air quality directly to characterize episode severity.

Note that if the episode is drawn from among the 3 years upon which the nonattainment designation is based, days which are chosen are likely to have monitored observations very close to the current design value. "Close to" could be defined in diagnostic tests in specific studies. In the absence of such information, we suggest "$\pm$ 10 ppb" as a default recommendation for purposes of prioritizing choice of episodes. If the base and current periods do not coincide, "close to" is within $\pm$ 10 ppb of the design value during the base period straddling the episode. If it is not feasible to meet this default criterion for all monitoring sites, meeting it at sites with *current* design values $\geq$ 85 ppb should receive greatest priority.

Choose days with intensive data bases. Preference should be given to days with measurements aloft, available measurements of indicator species (see Section 6.0) and/or precursor measurements. These preferences result from a desire to incorporate a rigorous model performance evaluation as a part of the attainment demonstration. This reduces the likelihood of "getting the right answer for the wrong reason". Thus, the likelihood of mischaracterizing ozone/precursor sensitivity is reduced.

Choose a sufficient number of days to enable the monitored attainment test to be based on several days at each monitoring site violating the NAAQS. Figure 3.3 indicates that the relative reduction factor computed at any given site appears to be robust if based on a mean response averaged over several days. Some studies imply that the relative reduction factor may be more variable if based on an individual or small number of days (Milanchus, et al., (1998)). An air quality model may also have greater success predicting 8-hour daily maxima matched in space if comparisons are based on means observed over several days. Therefore, States should model as many days as feasible.

We offer the following 7-step procedure as one which may be useful in combining the four primary criteria for selecting episodes to model.

1. For each episode being considered, States should examine *observed* 8-hour daily maximum concentrations at each (sites with design values < 75 ppb can be excluded) monitoring site during the year of the episode, as well as during the year before and the year after the episode. Thus, if one is examining days in a 1991 episode for suitability in the attainment test, severity of the candidate days should be assessed relative to 1990-92 observations at each selected site.

2. For each of the three years, rank the top ten 8-hour daily maxima observed at each of the monitoring sites selected in step 1.

3. Compute the average 1st high 8-hour daily maximum, the average 2nd high 8-hour daily maximum, etc down to the average 10th high 8-hour daily maximum for each selected monitor.

4. Note a range of concentrations which are $\pm$ 10 ppb of the average 4th highest value at each site.

5. Classify qualifying days from step 4 into meteorological regimes, using observed or computed wind fields as the primary criterion for classifying the regimes.

6. Note days in the preceding sample for which intensive data bases exist.

7. Give priority to choosing a mix of episodes containing days with observations ± 10 ppb of the site-specific design values during the base period(s), drawn from a variety of meteorological classes identified in step 5, and for which observations aloft, indicator species and/or precursor measurements are available. Try to choose a sufficient number of days so that several days are suitable for use in the modeled attainment test applied at each site violating the NAAQS.

Recommendations. States should consider four primary criteria when choosing meteorological episodes for modeling. Tradeoffs among these may often be necessary. Such tradeoffs need to be resolved on a case by case basis.

1. Choose frequently occurring episodes containing days reflecting a variety of wind orientations observed to occur when 8-hour daily maxima exceed 84 ppb at one or more monitors.

2. Choose episodes containing days with observed 8-hour daily maximum ozone concentrations close to (e.g., ± 10 ppb) the average 4th high daily maximum observed at monitoring sites during a 3-year period straddling the period from which each episode is drawn (i.e., days approximately as severe as implied by the form of the NAAQS).

3. Choose episodes containing days for which measurements aloft, measurements of indicator species and/or precursor measurements exist.

4. Choose a sufficient number of days so that several days are available for use in the modeled attainment test for each monitoring site where the NAAQS is violated.

12.2 What Additional, Secondary Criteria May Be Useful For Selecting Episodes?

In Section 12.1, we noted that there may often be conflicts among the 4 primary criteria recommended as the basis for choosing episodes to model. Several additional, secondary selection criteria may be helpful for resolving these conflicts.

Choose episodes which have already been modeled. That is, of course, provided that past model performance evaluation for such an episode was successful in showing that the model worked well in replicating observations. Given that the 4 primary criteria are met approximately as well by such episodes as they are by other candidate episodes, a State could likely save a substantial amount of work in evaluating model performance.

Choose episodes which are drawn from the period upon which the current design value is based. As we note in Section 3.3, fewer emission estimates and fewer air quality model simulations are needed if the "base period", used to evaluate model performance, and the "current period", used in the recommended modeled attainment test, are one in the same. Following this criterion could also make the second primary criterion more straightforward. That is, current air quality observations rather than episode severity estimated for a period several years ago could be used as a basis for choice of episodes. We discuss choice of a "current period" in Section 3.1. A "current period" may be either (a) the 3-year period which straddles the year of the most recent inventory (e.g., 1995-1997, when 1996 is the most recent inventory), or (b) the 3-year period used as the basis for the nonattainment designation (e.g., 1997-1999). We assume that the two choices very nearly coincide, so that if an episode comes from a period other than the one straddling the year of the inventory, needed inventory adjustments would be minor and can be readily made for performance evaluation and use in the attainment test.

Choose episodes having observed concentrations "close to" implied severity of the form of the NAAQS on as many days and at as many sites as possible. This criterion is related to the modeled attainment test and to the fourth primary criterion for episode selection. The more days and sites for which it is reasonable to apply the test, the greater the confidence possible in the modeled attainment test.

It is desirable to include weekend days among those chosen, especially if concentrations greater than 84 ppb are observed on weekends,. Weekend days reflect a different mix of emissions than occurs on weekdays. This could also lead to different spatial patterns of 8-hour daily maxima in excess of 84 ppb. Thus, for increased confidence that a control strategy is effective it needs to tested on weekends as well as on weekdays. If emissions and spatial patterns of high ozone do differ on weekends vs. weekdays, including weekend days in the choice of episodes may provide a potential for evaluating accuracy of a model's response to changes in emissions. As we note in Section 16.0, such evaluations are highly desirable.

If a State chooses to model several nonattainment areas simultaneously, choose episodes which meet the other criteria in as many of these nonattainment areas as possible. As discussed in Section 11.0, a State or group of States may decide to apply a regional model or a nested regional model to demonstrate attainment in several nonattainment areas at once. Time and resources needed for this effort could be reduced by choosing episodes which meet the other criteria in several nonattainment areas which are modeled.

<u>Recommendations</u>. States may be able to resolve conflicts among the primary criteria for selecting episodes by considering one or more secondary criteria. The following are identified as secondary criteria. States may identify, document and present the rationale for criteria in addition to these if they choose.

1. **Give preference to previously modeled episodes.**

2. Give preference to episodes occurring during the period corresponding to the current design value used in the modeled attainment test.

3. Give preference to episodes maximizing the number of days and sites observing 8-hour daily maxima close to the level of severity specified in the NAAQS.

4. Include weekends among the selected days, especially if daily maxima exceeding 84 ppb are observed on such days.

5. If applying a regional model, choose episodes meeting the other primary and secondary criteria in as many nonattainment areas as possible.

13.0 What Should I Consider When Selecting A Modeling Domain And Its Horizontal/Vertical Resolution?

A modeling domain identifies the geographical bounds of the area which is modeled. Recommended domain size depends on the nature of the strategies believed necessary to meet the air quality goal. This, in turn, depends on the degree to which air quality observations suggest that a significant part of an observed exceedance is attributable to regional concentrations which approach or exceed levels specified in the NAAQS. Choice of domain size is also affected by data base management considerations. Generally, these are less demanding for smaller domains.

Horizontal resolution is a function of the size of individual grid cells. Vertical resolution is determined by the number of grid cells (i.e., layers) considered in the vertical direction. Choice of horizontal grid cell size and a suitable number of vertical layers depends on spatial variability in emissions, spatial precision of available emissions data, mixing heights, likelihood that mesoscale or smaller scale meteorological phenomena will have a pronounced effect on precursor/ozone relationships, data base management constraints and computer/cost constraints.

We begin this Section by discussing factors States should consider in choosing domain size. Next, we address choice of horizontal grid cell size and number of vertical layers. We conclude by discussing factors affecting choice of size and resolution of coarse scale and fine scale grids considered in a nested modeling analysis.

13.1 How Do I Choose Between An Urban Scale Or Regional Domain?

States may find it useful to examine the gap between a nonattainment area's design value and the level specified in the NAAQS vs. the gap between observed regional (upwind) concentrations and the level specified in the NAAQS. If the former gap is less than the latter, an urban scale analysis may suffice. To illustrate for the case of ozone, if a nonattainment area had a design value of 95 ppb and regional 8-hour daily maxima were typically 60 ppb, the former gap (11 ppb) is substantially less than the latter gap (24 ppb). Depending on the judgment of those implementing the protocol, the strategy for meeting the NAAQS may thus focus on local control measures. An urban scale domain size may be appropriate. In contrast, if the local design value were 95 ppb but corresponding regional daily maxima were typically 80 ppb, the former gap remains 11 ppb, but the latter is reduced to 4 ppb. Those implementing the protocol may wish to consider using regional as well as local measures in such a case. This would necessitate using a regional modeling domain. In general, if additional regionally implemented control measures are expected to materially affect the amount of additional local controls needed to meet the air quality goal, a regional modeling domain should be used. If not, an urban scale domain should suffice.

What do we mean by "urban scale" and "regional" domains? An urban scale domain is one having horizontal dimensions less than ~ 300 km on a side. Assuming the nonattainment area is located near the center of the domain, the domain should be large enough to ensure that

emissions occurring shortly before sunrise in its center are still within the domain near the end of the same calendar day. If recirculation of the nonattainment area's previous day's emissions is believed to contribute to an observed problem, the urban scale domain should be large enough to characterize this. If recirculation encompasses distances larger than about 300 km, an urban scale model is probably not sufficient to address an area's problem.

A regional domain is one having horizontal dimensions typically exceeding 1000 km on a side. Data base management problems generally make it infeasible to use the same horizontal grid cell size in urban scale and regional models. Nested regional models are intended to address this problem. A nested regional model is one whose domain typically exceeds 1000 km on a side. However only a portion of that domain (e.g., < 300 km on a side) has grid cells with a size similar to that recommended for urban scale models. States should ordinarily include all monitoring sites considered in the modeled attainment test within the area covered by a grid with size of individual cells comparable to that recommended for urban scale modeling.

Recommendations. Selection of a domain size depends on the types of control strategies to be simulated. States should review regional (upwind) design values vs. those occurring in the nonattainment area to determine the emphasis to place on regional vs. local controls. If this review suggests that a regional strategy is an important component of an attainment demonstration, then the domain should be regional (>1000 km) in coverage. Otherwise an urban scale analysis (<~300 km) may suffice.

13.2 What Horizontal Grid Cell Size Is Necessary?

As we discuss in Section 14.0, we anticipate widespread use of dynamic meteorological models to provide meteorological inputs needed to make air quality estimates. The most commonly used of these models are set up to produce meteorological fields for 108, 36, 12 and 4 km cells. Thus, the issue addressed in this Section is which of these sizes to recommend as an upper limit for regional models and for urban scale or fine portions of nested regional grids.

In past guidance, we have recommended using horizontal grid cell sizes of 2-5 km in urban scale modeling analyses (U.S. EPA (1991)). Sensitivity tests performed by Kumar, et al. (1994) in the South Coast Air Basin compare hourly base case predictions obtain with 5 km vs. 10 km vs. 20 km grid cells. Results indicate that use of finer grid cells tends to accentuate highest hourly ozone predictions and increase localized effects of NOx titration during a given hour. However, statistical comparisons with observed hourly ozone data in this heavily monitored area appear comparable with the 5 and 20 km grid cells in this study. Comparisons between hourly ozone predictions obtained with 4 km vs. 12 km grid cells have also been made in an Atlanta study (Haney, et al. (1996)). As in Los Angeles, use of smaller (i.e., 4 km) grid cells leads to higher domain wide maximum hourly ozone concentrations. However, when reviewing concentrations at specific sites, Haney, et al. found that for some hours concentrations obtained with the 12 km grid cells were higher than those obtained with the 4 km cells. Since

signs of the differences obtained with 4 km vs. 12 km grid cells vary for different hours, this may suggest that 8-hour daily maximum ozone predictions are less sensitive to the selected grid cell size than 1-hour daily maxima. Recent sensitivity tests comparing relative reduction factors in predicted 8-hour daily maxima near 272 sites in the eastern United States indicate generally small unbiased differences ($\leq$ .04, in 95% of the comparisons) using a grid with 12 km vs. 4 km grid cells (LADCO (1999)).

Intuitively, one would expect to get more accurate results in urban applications with smaller grid cells (e.g., 4 km) provided the spatial details in the emissions and meteorological inputs support making such predictions. Thus, using 4 km grid cells for urban or fine portions of nested regional grids and 12 km cells in coarse portions of regional grids are desirable goals. However, extensive use of urban grids with 4 vs. 12 km grid cells and regional grids with 12 vs. 36 km grid cells greatly increases computer costs, running times and data base management needs. Further, elsewhere in this guidance we identify needs to model large domains, many days, and several emission control scenarios. We also identify a number of diagnostic tests which would be desirable and suggest using more vertical layers than has commonly been done in the past. Also, there may be means of dealing with potential problems posed by using larger than desired grid cells. For example, use of plume in grid algorithms for large point sources of NOx might be considered as an alternative with coarser than desired grid cells.

Relative importance of using a domain with grid cells as small as 4 km will need to be weighed on a case by case basis by those implementing the modeling/analysis protocol. Thus, in this guidance, we identify upper limits for horizontal grid cell size which may be larger than desired for some applications. This is intended to give States flexibility to consider competing factors (e.g., number of modeled days vs. grid cell size) in performing a modeling analysis within given limits of time and resources.

For coarse portions of regional grids, we recommend a grid cell size of 12 km if feasible, but in no event larger than 36 km. For urban and fine scale portions of nested regional grids, it may be desirable to use grid cells about 4 km, but, in no event larger than 12 km. All ozone monitor locations within a nonattainment area should ordinarily be placed within the fine scale portion of a nested regional grid if nested models are used. States choosing an urban grid or fine portion of a nested grid with cells larger than 5 km should undertake several additional analyses. First, States should apply plume in grid algorithms to major point sources of NOx if they choose an urban or fine portion of a regional grid with cells as large as 12 km. Once an emission control strategy has been tentatively selected, States should test the current and the selected control strategy with grid cells $\leq$ 5 km, if feasible, so that the outcome is available to be considered in a weight of evidence determination.

Recommendations. Horizontal grid cell size in regional models should be $\leq$ 36 km, except in areas used to establish boundary conditions for the regional model (where they may be larger). For urban scale analyses and the fine scale portion of a nested regional model, cells which are 4-5 km on a side are preferred, if feasible. Cells should not exceed 12 km on a side in these analyses. If cells as large as 12 km are used in

urban areas, States should consider using plume in grid algorithms to deal with large point sources of NOx. States should perform diagnostic sensitivity tests to see whether using grid cells smaller than 12 km affects conclusions reached in the modeled attainment test when the selected control strategy is simulated. If so, this should be considered in a weight of evidence determination.

13.3 How Many Vertical Layers Should I Consider?

As described in Section 14.0, the preferred means for generating meteorological data fields for input to air quality simulation models is to use a dynamic meteorological model with four dimensional data assimilation (FDDA). Such models often consider as many as 20-30 vertical layers. To minimize a number of assumptions needed to interface meteorological and air quality models, it is better to use identical vertical resolution in the air quality and meteorological models. However, application of air quality models with so many vertical layers may not be feasible nor cost effective. In this Section we identify factors affecting number of vertical layers chosen for use in an air quality model, as well as the placement of these layers.

Accuracy of predicted base case ozone concentrations depends, in part, on how accurately a model is able to characterize dilution of precursors and ozone. This, in turn, depends on how precisely the model can estimate maximum afternoon mixing heights (i.e., the planetary boundary layer). Precision of mixing height estimates is affected by the thickness of the model's vertical layers aloft which are near the anticipated mixing height (Dolwick, et al., (1999)). Because maximum mixing heights may vary on different days and it is necessary to simulate numerous days and locations, model predictions can be influenced by the number of vertical layers considered by the model. Dolwick, et al., (1999) have shown that base case predictions are not sensitive to the number of layers considered *above* the planetary boundary. Thus, States may assume as few as one layer above the highest conceivable maximum afternoon mixing height with the rest of the vertical layers occurring within the planetary boundary layer.

Placement of vertical layers within the planetary boundary layer is also an important issue. For practical reasons, it is best to have an air quality model's vertical layer placement coincide with layers considered in the meteorological model used to generate meteorological inputs. So the placement issue really is, which ones of the boundaries between the meteorological model's layers should one match with the boundaries between vertical layers used in the air quality model? Based on the discussion in the preceding paragraph, we recommend high precision near the anticipated maximum afternoon mixing height. In addition, observed 8-hour daily maximum ozone concentrations may well include some evening hours. Surface concentrations during these hours may be affected by presence of a low level inversion whose base is just above turbulence introduced by surface roughness or, in some cases, by an urban heat island. Thus, States should use a shallow surface layer, generally no more than 50 meters in depth. In general, layers below the mixing height should not be too thick, or large, unrealistic step increases in mixing may occur. States should try to avoid using layers within the planetary

boundary layer thicker than about 200-300 meters.

Based on recent sensitivity studies (Dolwick, et al. (1999) and LADCO (1999)), it appears as though as few as 7-9 vertical layers (including one above the planetary boundary layer) may suffice in a modeling study if care is taken in specifying placement of these layers. Prior to modeling, we recommend that States review available meteorological measurements aloft to get a sense of where the maximum afternoon mixing height is likely to be on days which might be modeled. We recommend that the number of vertical layers considered in coarse and fine portions of a nested regional grid be identical.

Recommendations. An air quality model may consider fewer vertical layers than commonly considered in a meteorological model. However, boundaries between vertical layers used in an air quality model should coincide with selected boundaries in the meteorological model. Care should be taken to configure the vertical layers so that the maximum afternoon mixing height is defined as precisely as possible. The surface layer considered in the model should generally be no more than 50 meters deep, and no layer beneath the mixing height should be more than about 300 meters thick. The minimum number of layers chosen depends on the meteorological conditions and characteristics of the area to be simulated. To meet the preceding criteria, States should generally use at least 7-9 vertical layers within the planetary boundary layer and 1-2 layers above it.

13.4 What Else Should I Consider In Choosing Finely And Coarsely Resolved Portions Of Nested Regional Models?

Coarse Grid Domain. Size of a coarse grid domain should be consistent with the chemical/physical lifetimes of pollutants to be modeled. It should also reflect the purpose for which regional modeling is undertaken. For example, if a regional analysis is performed to assess effects of a regional strategy simultaneously for a number of nonattainment areas, the domain needs to be larger than if only a limited number of nearby areas were the focus of the study.

Lifetimes vary for ozone and its precursors. Lifetime for NOx (i.e., NO + NO_2) may be less than a day. Regional analyses performed in the U.S. to date suggest that lifetimes for sulfates and nitrates are two days or less (Dennis, 1994). Sources of VOC are believed to be ubiquitous, due to natural emissions. Many of these natural emissions are relatively reactive, so that multi day transport of stable species of VOC or radical products resulting from oxidation of more reactive species may not be a critical factor for selecting size of a domain for modeling ozone. Lifetime for ozone is notoriously difficult to estimate due to the recycling of this compound with free radicals, concentrations of oxidized species of nitrogen and emissions of fresh NOx and VOC precursors which occur in transit. Given information about the lifetime of nitrates however, it is probably safe to assume a lifetime for ozone which is on the order of 2-3

days. The foregoing information suggests that, ideally, the size of a regional modeling domain should be large enough so that emissions occurring two days prior to the beginning of daylight on a modeled day of interest are included within the domain. Thus, we suggest States focus on their receptor areas of interest, perform some screening analyses with trajectory models to ensure that major source areas within two days' travel time are included in the domain.

Fine Grid Domain. Size of the fine grid domain should be influenced by several factors: (1) proximity of receptor sites to major sources of ozone precursors (especially NOx); (2) presence of topographical features which appear to affect observed air quality; and (3) desire to limit resource intensive efforts needed to use numerical models on a fine scale. The last factor is an important concern for use of nested regional models. Size of a fine grid domain could be smaller than that recommended for an urban scale analysis. This follows, since the coarse domain is available to estimate impacts of sources located at intermediate distances from the receptor area, whereas this information is not available for an isolated urban scale analysis. The issue of how far to extend a fine scale grid is one which may need to be resolved on a case by case basis. We recommend that States examine the issue using diagnostic sensitivity tests (see Section 16.0). For consistency with the modeled attainment test, we recommend that the fine grid should initially extend 15 km (i.e., 3-4 4-km grid cells) beyond any receptor of interest.

<u>Recommendations</u>. Size of a coarse grid should be large enough to include potentially important sources located two days' travel time from receptor sites of interest. Applications which need to consider numerous receptor sites located some distance apart therefore need to use larger domains than do applications focusing on receptors in close proximity to one another. Extent of a fine grid also depends on the number of receptor sites. States should perform diagnostic analyses to ascertain how far a finely resolved grid needs to extend. As a starting assumption, we recommend extending the finely resolved grid sufficiently so that it extends at least 15 km beyond all monitoring sites considered in the modeled attainment test.

14.0 How Do I Produce Meteorological and Air Quality Inputs Needed By An Air Quality Model?

After episodes are selected for modeling, corresponding meteorological inputs need to be generated for use in an air quality model. Although the resulting inputs remain constant, they can affect outcomes of a number of the modeling outputs we have identified for scrutiny in Section 4.1. They may also potentially affect relative reduction factors used in the attainment and screening tests. In contrast to meteorological data, air quality inputs may change between times corresponding to "current" and "future" emissions used in the modeled attainment test. This presents a potential problem which needs to be addressed.

In this Section, we describe two approaches for generating meteorological inputs to air quality models, and identify advantages/disadvantages associated with each. We note that using dynamic meteorological models with output "nudged" by observations is usually the preferred approach for generating needed meteorological data. For some applications, use of these models for horizontal grid cells smaller than 12 km may present practical problems. We identify ways to diminish these, if they occur. It is important to quality assure meteorological inputs prior to their being used in an air quality model. We next discuss how these inputs can be evaluated. We conclude by identifying the role of air quality inputs as initial and boundary conditions for an air quality simulation, and note ways to reduce limitations to the simulation resulting from sparseness of these data.

14.1 What Approaches Are Available For Generating Meteorological Data?

Two approaches have been widely used to generate meteorological data needed in air quality models for ozone. The first of these (diagnostic models) relies primarily on observed data and introduces some additional constraints on wind flow due to terrain features. Observed surface temperatures and sounding data are used to develop other information needed to characterize mixing.

Most frequently used diagnostic wind models are described by Douglas, et al. (1990) and by Scire, et al., (1998). The main advantage of diagnostic models is that they are relatively easy and inexpensive to apply. Further, they make maximum use of wind observations. There are several disadvantages, however. First, there are seldom enough observations to adequately define a windfield, particularly aloft. Much of the input to the air quality model is derived through interpolation or subjective methods. Because of the sparseness of observations in many areas, we do not encourage use of diagnostic models for generating inputs to regional scale air quality model applications. A second disadvantage is that the meteorological estimates derived with a diagnostic model are not necessarily physically consistent with one another. In the atmosphere, there is a physical dependency existing between temperature, pressure and windfields. This interdependency is not extensively accounted for in diagnostic models, and the extent to which it is considered depends on the expertise of those applying the model. Nevertheless, if ambient concentrations of a pollutant (e.g., as for one or more components of

$PM_{2.5}$) are believed to be primarily affected by winds and urban scale source/receptor orientation, the disadvantages are not serious enough to preclude use of diagnostic models.

The second approach for generating needed meteorological data is to use dynamic meteorological models with four dimensional data assimilation (FDDA). These models attempt to characterize theoretical relationships between meteorological variables and topographical/terrain characteristics. Use is made of relatively sparse observations aloft to help steer (i.e., "nudge") solutions so that they do not diverge from observed meteorological fields. Wind observations aloft are typically used for this purpose. See Seaman, (1997) for a further summary of the attributes of dynamic meteorological models. The MM5 (Grell, et al., (1994) and Seaman, et al., (1996)), RAMS (Pielke, et al., (1992) and Lyons, et al., (1995)) and the SAIMM (Systems Applications International, (1996)) models are among those which have been most widely used with numerical air quality models. The major advantage of dynamic meteorological models is that they provide a way of characterizing meteorological conditions consistent with theory, terrain and each other at times and locations where observations do not exist. Disadvantages have been large required computer resources and considerable expertise needed to apply the approach. Recent advances in computer technology have resulted in increased use of dynamic meteorological models for air pollution applications. The MM5 model is used as the default approach with the CMAQ model in MODELS3. States need to consider compatibility between candidate meteorological models and the air quality model(s) chosen for use. We believe that use of dynamic meteorological models with FDDA is generally the preferable approach for generating meteorological inputs to air quality models for ozone.

Although improvements in computers have made increased use of dynamic meteorological models possible, we have found that data storage requirements and CPU time increase dramatically as the horizontal grid cell size required of the meteorological model becomes finer. For example, the CPU time needed to generate meteorological data resolved to 12 x 12 km grid cells is considerably greater than the expected factor of "9" increase in that needed to process meteorology for a domain with 36 x 36 km grid cells. This suggests that States may need to limit the spatial extent of areas and the number of episodes for which dynamic meteorological models are used to process meteorological data for grids with horizontal resolution <12 km. Generally, a finely resolved meteorological field needs to extend about 3 grid cells beyond the bounds of the fine scale grid used to make air quality predictions. For example, if 4 km grid cells were used in the fine portion of a nested regional air quality model, meteorological fields at this detail would need to extend 12 km beyond the bounds of the 4 km grid used for air quality predictions.

Recommendations. States should ordinarily use a peer reviewed dynamic meteorological model with four dimensional data assimilation as the means for generating meteorological inputs to ozone models. Peer reviewed diagnostic models may be used on a case by case basis. Grid cell size used in dynamic models should be chosen considering factors discussed in Section 13.0.

14.2 How Do I Deal With Data Management And Computer-related Constraints When Applying Dynamic Meteorological Models?

States should ordinarily use dynamic meteorological models resolved to the same level as desired for making air quality predictions. Occasionally, this may not be feasible, or may lead to poor performance of the dynamic model. In this Section, we identify possible means for reducing one or both of these problems. The methods we discuss may increase the risk of discontinuities at the bounds of a finely resolved grid. These should be checked and corrected to the extent possible before proceeding.

The first approach is to use available results from dynamic models on the next greatest coarse scale (i.e., 36 km for a desired 12 km estimate, 12 km for a desired 4 km estimate) to interpolate more finely resolved fields. An objective approach like bilinear interpolation could be used (U.S. EPA, 1991). This approach would be particularly useful if the major reason for desiring finely resolved meteorological estimates is related to a need to resolve emission estimates more finely. For example, in the case of ozone, fine grid cells may be needed to most accurately characterize the apparent detrimental effect of NOx reductions on predicted ozone resulting from titration of ozone by nitric oxide near sources of NOx.

A second approach for circumventing major resource requirements needed to apply dynamic models for finely resolved grids considers topographic information (e.g., presence of land/water interfaces) and measured meteorological data to refine fields coarser fields generated by a dynamic model. This second approach may be preferred if the major reason for desiring finely resolved meteorological inputs has to do with perceived importance of mesoscale features which cannot be adequately considered through an objective interpolation procedure. In essence, the second approach is to apply a diagnostic wind model to the wind field generated by the more coarsely resolved dynamic model.

Finally, consequences of using coarse grid cells (e.g., 12 km when 4 km might be more desirable) can be reduced by specifying a land use for each cell that corresponds to usage near the major portion of emissions within a cell. This approach is most applicable at land/water interfaces. By assuming the cell is entirely "land", vertical dispersion of fresh emissions is likely to be better characterized. This might also result in a better characterization of subsequent transport of coastal emissions over adjacent large bodies of water.

Recommendations. Prohibitive computer-related constraints associated with applying a dynamic meteorological model to derive a finely resolved (4-12 km) set of meteorological data can be addressed in one of two ways.

1. Interpolate more coarsely resolved data using objective analysis.

2. Apply a diagnostic wind model using "observations" generated by the dynamic

meteorological model for a coarser grid. Assume other variables remain the same as for the coarser grid.

Consequences of using coarser than desired grid cells may be reduced by assigning a land use factor for each surface cell which corresponds to the location of most emissions within the cell (e.g., at cells including an interface between land and a large body of water).

14.3 How Do I Quality Assure Results Generated By A Meteorological Model?

There are several ways to evaluate performance of a meteorological model. Although it is desirable to evaluate meteorological inputs *before* air quality predictions are made, some of the means available for evaluating the meteorological model's predictions must wait until the air quality model is run. Important meteorological outputs warranting scrutiny include wind velocity patterns, mixing heights (e.g., estimated by noting the vertical layer at which vertical diffusivity (K_v) is suppressed), temperature, pressure, water vapor and cloud cover. Methods for evaluating output from a meteorological model include comparison with selected upper air measurements, derivation of trajectories, use of computer graphics, use of non-reactive tracers, comparing results obtained with different models, use of dimensionless parameters, comparing spatial patterns of observed and predicted daily maximum ozone and use of process analysis. Each of these is briefly described in the following paragraphs.

Comparison with upper air observations. This can be done by excluding selected upper air observations from use in four dimensional data assimilation (FDDA) so that they can be used to assess model performance. Wind velocity, temperature, pressure and water vapor are important variables to compare. If aloft measurements are available at more than one altitude, they can provide a means for evaluating how well a model characterizes vertical exchange in the lowest few layers. Generally, routine data bases (i.e., widely separated soundings taken twice per day) are needed to support FDDA. The data base is probably insufficient to exclude data to evaluate model performance. In Section 5.0, we noted that it is desirable to increase measurements aloft. One reason for doing this is to provide better means for evaluating performance of meteorological models.

Derivation of trajectories. A State could select several locations in the grid and use trajectory models such as HY-SPLIT (NOAA, 1999) to derive back- or forward-trajectories from the hourly wind fields generated by a meteorological model. If surface trajectories were limited to daylight hours, the computed trajectory could be compared with observed surface air quality observations. If the timing of high ozone observed along the path of the trajectories is consistent with expectations, given the configuration of sources, this would be an indicator that the meteorological model is performing adequately. A State could also derive daytime surface trajectories using *observed* wind data. These trajectories could also be compared with air quality patterns. By comparing the two sets of trajectories with observed air quality patterns, it would

then be possible to assess whether the meteorological model increases the skill with which ozone plumes are oriented.

Use of computer graphics. Examining wind vectors for apparent discontinuities is possible using graphics. It is also possible to construct difference diagrams between observed and predicted temperatures and winds. Locations where agreement is poor may suggest areas needing more finely resolved estimates. Geographical orientation between areas of poor agreement and locations of major sources or observed poor air quality may be plotted to judge potential significance of any disagreement.

Simulation of inert tracers. This approach is to assume a uniform concentration field (e.g., 10 ppb) of an inert tracer in an air quality model with grid cells identical to the horizontal and vertical size of the cells used in the meteorological model (this may be feasible since it is unnecessary to consider atmospheric chemistry, deposition or fresh emissions). Identical, constant boundary conditions should also be assumed. In theory, the concentration field should remain uniform, and there should be no systematic drift in the mass of material remaining within the grid. Predicted concentrations of the tracer can then be examined to see whether there are major discontinuities in the concentration field or problems with mass balance. If there are, this may suggest a problem with the meteorological model or with the ability of the air quality model to consider divergence/convergence predicted with the meteorological model.

Compare results obtained with different models. This approach is to compare results from two different models for a subset of days being considered. For example, MM5 and RAMS results could be compared to note differences in predicted surface temperatures as well as wind velocities at the surface and aloft. Reasons for major differences would then need to be diagnosed.

Compare estimated divergence or dimensionless parameters with expected ranges. Calculations can be made in selected portions of the grid to see whether they appear reasonable.

Compare spatial patterns of air quality predicted with a photochemical grid model with observed patterns on the days of interest. If the predictions are systematically skewed from the observations, this could suggest a problem with the meteorological outputs generated by the meteorological model.

Use process analysis. Process analysis applies to the output generated by an air quality model. It is described by Jeffries, (1997) and by Lo, et al., (1997). Its use with air quality models is noted in Section 16.0. Process analysis determines the relative importance of different chemical or physical factors as contributors to predicted ozone concentrations. If process analysis suggests that a variable influenced by meteorological inputs, such as vertical exchange (i.e., vertical diffusivity), plays a large, unanticipated role leading to a high ozone prediction, this might warrant a closer examination of what led to such a prediction.

Recommendations. To the extent possible, States should quality assure results from meteorological models prior to using them in the intended air quality model. States should select a mix of approaches for evaluating meteorological inputs to an air quality model on a case by case basis. Candidate approaches include:

1. **comparison with upper air measurements "held back" from use in FDDA;**
2. **comparison of calculated trajectories with observed air quality patterns;**
3. **use of computer graphics to discern spatial discrepancies;**
4. **simulation of inert tracers to identify discontinuities or mass balance problems;**
5. **comparing results obtained with different meteorological models;**
6. **calculating and comparing divergence and/or dimensionless parameters and comparing these with expected ranges;**
7. **comparing spatial ozone patterns obtained with a grid model vs. observed patterns, and**
8. **using process analysis to flag contributions made to unexpected ozone concentrations by meteorological factors.**

14.4 What Are Some Past Applications Of Dynamic Meteorological Models?

Table 14.1 lists some recent air quality modeling applications using the two most widely available dynamic meteorological models. Choice of a meteorological model may be influenced by compatibility with a chosen air quality model, as well as by past experience of those applying the air quality model. The listing in Table 14.1 is not comprehensive. Inclusion on the list does not necessarily imply an endorsement for a specific application. Exclusion does not necessarily imply that an approach is inappropriate for a specific application. States should consider using methods such as those in Section 14.3 to determine whether the output generated by a meteorological model is adequate for use in a specific application.

14.5 How Do I Address An Air Quality Model's Need For Air Quality Inputs?

Air quality inputs are needed in air quality models for two purposes: to specify initial conditions, and to specify boundary conditions. There is no satisfactory way to specify initial conditions in every grid cell. Thus, we recommend beginning a simulation at least one day prior to a period of interest for urban scale applications, and two days prior to periods of interest for

regional applications to diminish importance of arbitrary assumptions about initial conditions.

Table 14.1. Some Past Applications of Dynamic Meteorological Models

Meteorological Model	References Describing Model Performance	Sponsors (Applications)
MM5	Seaman, et al., (1995), (1996b) Tesche, et al., (1993a), (1993b)	CARB (San Joaquin Valley, South Coast Air Basin, CA)
RAMS	Tesche, et al., (1993c), (1993d), (1993e), (1993f)	LADCO (eastern U.S., with emphasis on Lake Michigan States)

Boundary conditions can be specified in several ways. One way is to nest the area of interest within a much larger domain. This approach is exemplified by using nested regional models, as described previously. The need to diminish importance of boundary conditions is why we recommended in Section 13.0 that States use a large regional domain, with the upwind bound 2 or more days' travel time from the area(s) which is the focus of an analysis. If it is not practical to use a nested regional modeling approach or if boundary conditions are believed to be relatively unimportant, a second approach is to use a large single domain in an urban scale analysis. The domain should be approximately symmetrical about the major local sources affecting local monitoring sites of interest, and should be large enough so that emissions occurring in the center of the domain just before sunrise remain within the domain until the end of the same calendar day. If recirculation is thought to be part of the problem, the domain size would need to be extended to be able to consider it. Use of a large, single domain requires one to make use of monitored data and interpolation to estimate boundary conditions. This approach begs the question about what to assume for *future* boundary conditions. It works best where boundary conditions are low and are expected to remain so.

Recommendations. Simulations should begin at least one day prior to the period of interest for urban applications and two days for regional applications. Use of nested regional models is the preferred approach for addressing boundary conditions. Where such an approach is not feasible, States should consider a single domain large enough to ensure that emissions occurring in the center of the domain just before sunrise remain within the domain until the end of the same calendar day or that next-day recirculation (if important) can be considered.

(This page is left blank intentionally)

15.0 How Do I Produce Emission Inputs Needed For An Air Quality Model?

Developing emissions inputs needed in air quality models requires several steps. First, States need to compile statewide and countywide emission estimates for precursors of ozone, as well as information subsequently used to spatially and temporally allocate emissions within each county included in the modeling domain. The most recent commonly available emissions estimates should be used in the modeled attainment and screening tests, described in Section 3.0. Next, an emissions model is used to convert the countywide emission information into hourly, gridded, speciated estimates needed by the air quality model. The emissions model also makes use of meteorological information (e.g., temperatures) to adjust emissions from some sources whose emissions are affected by environmental conditions. Both of these steps need to be accompanied by continual efforts to quality assure estimates as they are being made. The final emissions input needed in the modeling process is gridded, speciated, hourly emissions estimates which have been projected to a future year. The projections reflect net effects of growth and the control strategy which is to be simulated with the air quality model.

In the following Sections, we identify information which needs to be compiled on a state- and countywide basis. We then identify emissions models which can be used to convert countywide estimates to the inputs needed by air quality models. We next describe several approaches useful for quality assuring estimates obtained as the first two steps proceed. We conclude with a short discussion of emission projection methods.

While our discussions focus on air quality model needs, there is more extensive guidance available on emissions inventory requirements and preparation (U.S. EPA (1999c)). Additional guidance has been developed to prepare emission inventories. Much of this guidance for has resulted from a joint State/local agency and U.S. EPA effort called the Emission Inventory Improvement Program (EIIP). A series of seven documents has been issued as a result of the EIIP:

- Volume I: Introduction and Use of EIIP Guidance for Emissions Inventory Development (U.S. EPA, (1997a))
- Volume II: Point Sources Preferred and Alternative Methods (U.S. EPA, (1997b))
- Volume III: Area Sources Preferred and Alternative Methods (U.S. EPA, (1997c))
- Volume IV: Mobile Sources preferred and Alternative Methods (U.S. EPA, (1997d))
- Volume V: Biogenics Sources Preferred and Alternative Methods (U.S. EPA, (1997e))
- Volume VI: Quality Assurance Procedures (U.S. EPA, (1997f))
- Volume VII: Data Management Procedures (U.S. EPA, (1997g))

In addition, guidance exists or is being prepared on emission projections, the National Emission

Trends inventory methodology, and temporal allocations, spatial allocations, and chemical speciation of emission inventories (U.S. EPA, 1998d, 1998e, 1998f). The EIIP documents are available electronically through the U.S. EPA Internet website at ***http://www.epa.gov/ttn/chief/eiip/techrep.htm.*** States should consult these documents as they prepare their emission inventories.

15.1 What Countywide Emission Estimates Are Needed To Support Air Quality Models?

Statewide and countywide emissions need to be divided into 4 broad categories: stationary point source emissions, stationary area source emissions, mobile emissions for on-road and off-road sources and biogenic/geogenic emissions. Point sources should be classified by SCC and have associated location information (e.g., latitude/longitude coordinates) as well as diurnal and weekly operating schedules. Area source emissions should be classified by SCC and reported by county. Surrogate factors, used to spatially allocate emissions from the source category within an air quality model grid superimposed over the county, should be identified for each area source category. Defaults for surrogates are available in current emissions models. Examples of surrogate factors might be such things as population or employment by census tract, land use, etc. If information exists concerning weekly and diurnal emission patterns for different area source categories, this information should also accompany the state- and countywide area source emission estimates. On-road and off-road mobile source emissions should be estimated using the most current version of the U.S. EPA MOBILE model (or, in California, the current version of EMFAC) in concert with activity (i.e., vehicle miles traveled (VMT)) estimates. The mobile source emission estimates should be accompanied by recommended surrogates for spatially disaggregating the mobile emissions and by diurnal and weekly activity patterns so that gridded, hourly estimates can be obtained for mobile emission estimates in subsequent steps. We recommend States distinguish weekend vs. weekday activity levels for mobile and stationary area sources. Estimates for biogenic emissions can be made using the BEIS2 emissions model (Geron, et al., 1994) or updates approved by the U.S. EPA. A State should report biogenic emissions on a county basis. Information regarding spatial pattern of land use is needed within each county if a State wishes to distribute biogenic emissions within a county in a non-uniform manner.

For model applications addressing the ozone NAAQS, emission estimates for each source category should include countywide VOC, NOx and CO estimates for each month of the year. The VOC estimates should be accompanied by a recommended speciation profile for each source category. We recommend that States rely on local measurements to the maximum extent possible for the speciation profile estimates. However, default information on VOC species profiles is available in U.S. EPA (1993), if needed. These data and updates can be obtained electronically through the U.S. EPA's Internet website at ***www.epa.gov/ttn/chief/software.html#speciate.***

Recommendations. States should be familiar with guidance in U.S. EPA (1999c) and with U.S. EPA Emission Inventory Improvement Program guidance describing appropriate procedures for estimating statewide and countywide emissions needed to support SIP revisions for ozone. Air quality models require emission estimates from point, area, mobile and biogenic sources. In order to convert this information for use in air quality models, VOC species profiles, rationale for suballocating emissions within a county and for assuming diurnal and weekday vs. weekend variability in emissions is needed for each point source and for each major area source category, as well as for mobile sources. Default assumptions for spatial/temporal emission allocations are available in emissions models. However, assumptions which are more appropriate for a specific area can be substituted for these. Emission estimates for VOC, NOx and CO are needed for each month of the year to support possible use of regional model applications performed for warm weather cities and future needs to integrate ozone and $PM_{2.5}$ control strategies.

15.2 Can I Use the National Emissions Trends Inventory As A Starting Point?

If there are no previously available modeled inventories to serve as a starting point, we recommend that States derive an inventory suitable for use with models starting from the National Emissions Trends inventory (NET) (U.S. EPA, 1998d). The most recent NET reflects statewide, annual emission estimates for VOC, NOx, and CO for 1996. However, the U.S. EPA plans to have a 1999 NET available during the latter half of 2000. If available on a timely basis, the 1999 NET is the preferred starting point for estimating emissions needed to support modeling underlying the 2003 SIP revision. The EPA inventory guidance (U.S. EPA, 1999c) allows States to select any year from 1996 to 1999, although use of 1999 is encouraged. If the NET is used, it should be for the same year as the current inventory used by the State. Statewide emissions, by county, are in the NET and are available electronically through the U.S. EPA Internet website at ***www.epa.gov/oar/oaqps/efig/ei#Trends***. If a State is performing a regional or nested regional modeling analysis, the NET can often serve to provide countywide estimates for locations far removed from the area which is the focus of the modeled attainment demonstration. Closer in, States should quality assure and improve emission estimates as necessary. The NET may be used, at a State's discretion, where there have been no previous State-sponsored efforts to compile inventories.

15.3 How Do I Convert Countywide Inventory Information Into Data Used In Air Quality Models?

Air quality models predicting ozone require day specific hourly emission estimates for VOC, NOx and CO for each cell of a grid superimposed over the area modeled. Typically, there are thousands of grid cells in a model application. To utilize atmospheric chemistry in the air quality simulation model, VOC emissions also need to have their component chemical species identified. We recommend that source specific, local information be used for this purpose

whenever possible. The U.S. EPA maintains the SPECIATE data base. SPECIATE can be used when more source-specific information is lacking. It may be accessed electronically at ***www.epa.gov/ttn/chief/software.html#speciate***. Finally, emission factors for some sources are dependent on meteorological conditions such as temperature. Thus, meteorological conditions need to be known to estimate day specific emissions. Emissions models should be used to account for the numerous and diverse factors which need to be considered to derive emissions inputs to air quality models. Currently, separate models are used to prepare estimates from anthropogenic stationary vs. mobile sources and from biogenic sources.

Anthropogenic emissions from stationary sources. Two emissions models have been widely used to convert estimated emissions from stationary sources for use in air quality simulation models for ozone-related applications.. The first of these, EPS2, has been frequently used in past urban scale modeling applications for ozone, but more recently has also been used in regional applications (Causley, et al., (1990), U.S. EPA (1993a)). EMS95 is the second emissions model which has had wide use (Alpine Geophysics, Inc., 1995). EMS95 has been used in the modeling underlying the U.S. EPA's rule to reduce regional NOx emissions (U.S. EPA 1997h), as well as in other applications of nested regional air quality models.

The version of EPS2 described in Causley, et al., 1990 may be used only for urban scale model applications for ozone. However, there is an updated version which has been used in applications of a regional model for particulate matter and in a regional analysis for ozone performed in the Gulf States (U.S. EPA, (1993a)). A newer emissions model, Sparse Matrix Operator Kernel Emissions (SMOKE), has had limited use to date (MCNC, 1999). SMOKE is similar theoretically to EMS95. However, it is computationally more efficient, reducing time and memory required to formulate individual control strategies simulated in an air quality model.

Anthropogenic emissions from mobile sources. MOBILE5A is the most current available model to estimate mobile emission factors for ozone precursors from a vehicle fleet representative of any specified year (U.S. EPA, 1994a). The U.S. EPA's Office of Mobile Sources (OMS) is developing the MOBILE6 model for highway vehicles as well as a NONROAD model to improve estimates for off-highway vehicles. These two models are expected to be available by the end of 1999. Estimated emissions obtained with the new models may differ from estimates obtained with currently available models. States may track the status of MOBILE6 and NONROAD at the following internet addresses: ***http://www.epa.gov/omswww/m6.htm*** (MOBILE6) and ***http://www.epa.gov/oms/nonrdmdl.htm*** (NONROAD model).

Prior to the availability of MOBILE6 and NONROAD, States other than California should use MOBILE5A or any update to this model identified as appropriate by the U.S. EPA's Office of Mobile Sources for highway and off-highway vehicles. The website ***http://www.epa.gov/omswww/models.htm*** is a useful source of information on MOBILE5a and mobile source models in general. Resulting emission factors need to be combined with activity levels (e.g., vehicle miles traveled) to estimate emission levels which have been suitably

disaggregated spatially and temporally for use as inputs in air quality models. Methods for estimating activity levels are included in U.S. EPA, (1997d).

Biogenic Emissions. The BEIS2 emissions model is the most widely used procedure for estimating biogenic emissions (Geron, et al. 1994 and U.S. EPA, 1997e). This model requires a mix of land uses to be specified for each county, as well as hourly temperature information. If a State believes the average land use mix characterized for a county is inappropriate for certain gridded locations within a county, this may be overridden for the grid cells in question on a case by case basis. The model makes use of stored information regarding geographic distribution of plant species, as well as the provided land use and temperature information, to generate gridded biogenic emissions.

Table 15.1 summarizes available emissions models used in recent air quality model applications, and identifies some example applications.

Table 15.1. Some Emissions Models And Example Applications

Emissions Model	References	Sponsors (Applications)
EMS95	Alpine Geophysics, (1995)	LADCO (eastern half of the U.S.), U.S. EPA, OAQPS (eastern half of the U.S.), NY DEC (eastern half of the U.S.).
EPS2	U.S. EPA (1993a).	U.S. EPA, Region IV (Gulf States), U.S. EPA, OAQPS (nationwide)
SMOKE	MCNC, (1999)	NC DEM (Charlotte, most of NC and parts of surrounding States)
MOBILE or EMFAC with activity estimates	U.S. EPA, (1997d)	MOBILE: Many sponsors (throughout the U.S. outside of California) EMFAC: CARB (California)
BEIS2	Geron, et al., (1994), U.S. EPA, (1997e)	U.S. EPA OAQPS (eastern half of the U.S.)

Recommendations. States should use emissions models to convert emission inventory estimates into emissions inputs required by air quality models. Emission models require additional inputs concerning chemical speciation, spatial and temporal disaggregation. Choice of models depends on compatibility with the chosen air quality model and the application at hand, as well as past experience of those implementing the modeling/analysis protocol. States should quality assure outputs from emissions models prior to making air quality estimates.

15.4 What Should I Do To Quality Assure Emissions Estimates?

The most efficient means to quality assure (QA) emission estimates is to apply QA during the initial emissions estimation process. The previously mentioned EIIP quality assurance document, U.S. EPA (1997f), contains a number of QA procedures that should be used to develop the basic countywide emission inventory. Once the basic emission inventory is ready for modeling, there are three additional quality assurance techniques that may be appropriate. The first is to compare emission estimates to estimates reported elsewhere. States can use results from such comparisons to see whether their estimates are unusual in any way. This focuses attention on portions of the inventory which appear to differ from estimates made for other locations, so that a State can confirm whether or not its initial estimates are appropriate. The NET inventory provided by the U.S. EPA may be useful for this approach.

Displaying emissions estimates graphically is also a useful means for quality assuring them. Emissions models identified in Section 15.3 can produce graphic displays useful for quality assurance. For example, a tile plot of emissions made for a grid superimposed over the area to be modeled is an effective means for identifying misplaced sources and for assuring oneself that spatial patterns of emissions are consistent with where sources are believed to be. Other graphical displays include pie charts and time series plots. Pie charts are useful for assessing whether distribution of emissions among source types or categories is consistent with expectations. Time series displays allow a State to look at estimated diurnal patterns in emissions to see whether these appear logical. They enable comparisons to be made for weekends vs. weekdays to see whether estimated differences appear reasonable.

Comparing emissions with monitored air quality is another means for quality assuring emissions estimates. As we place increased emphasis on measurements of ozone precursor species, comparison with monitored speciated data may become an increasingly important means for quality assuring emissions estimates. Availability of speciated VOC data, such as those in the PAMS network or similar data, makes it possible to use monitored observations to apply source attribution approaches (i.e., "receptor models"). A finding suggesting that air quality observations are the product of a mix of emissions which differs greatly from that inferred from

the inventory can point the way toward parts of the inventory which may need greater scrutiny. Receptor models and their uses have been summarized by Watson (1997) as well as in Seigneur, et al. (1997). Use of ambient data from the PAMS network to quality assure emissions estimates is described in U.S. EPA (1996a).

Recommendations. Quality assurance of emissions estimates is an essential part of the modeling process, and should be performed on a continual, ongoing basis. States should consider the following approaches to quality assurance: primary QA emphasis during the initial development of the basic emission inventory, comparison with available emissions estimates performed by others, computer graphics depicting emissions model estimates, and comparison with speciated air quality data.

15.5 How Do I Estimate Emissions For Future Years?

Emissions projections for sources within a modeling domain are needed to determine if a nonattainment area will meet the ozone NAAQS by the required future attainment date. For ozone, we require States to estimate future precursor emissions for at least one future date--two years before the date of required attainment. As discussed in Section 5.0, when a required attainment date is distant (e.g., 2010) from the date required for a SIP submittal (e.g., 2003), a State may wish to consider projecting emissions to an intermediate period (e.g., 2005-2007) as well.

The goal in making projections is to try to account for as many of the important variables that affect future year emissions as possible. Each State is encouraged to incorporate in its analysis the variables that have historically been shown to affect its economy and emissions the most, as well as the changes that are expected to take place over the next 10 to 20 years.

Each State should examine the source types that currently dominate its inventory and each should perform some rough calculations to see if that source distribution is likely to change much in the near future. This should suggest the emphasis that might be placed on projection methods for predominant source categories (if there are any). There is normally a wide range of ozone-precursor-emitting source types. Thus, it is probably only in exceptional cases where there are one or two major source types that dominate the inventory. Large point-source emitters in ozone nonattainment areas are already subject to Reasonably Available Control Technology (RACT) requirements and, in some cases, control technique guidelines (CTGs), which may be identical. Therefore, there may be many different emitters in ozone nonattainment areas whose emissions need to be tracked with time. In cases where there are a few dominant sources, special techniques should be used to ensure that those sources are modeled using more sophisticated techniques than those used for the rest of the inventory.

A State's needs for inputs to a grid-based model are a factor in making projections. As noted previously, grid-based models require source locations (coordinates) as input. Thus, a

projection approach that makes its computations at this level is preferred. A less desirable alternative is to assume that all growth and retirement occurs at existing facilities and that there is no variation in growth or control within each source category.

Information detailing the different types of projections that might be required of a State or local air pollution control agency can be found in the EPA publication "Procedures For Preparing Emissions Projections" (U.S. EPA, 1991a). In addition to the necessary types of projections, methods for projecting changes in future air pollution generating activities, quantifying the effects of current and future controls, and combining effects of growth and control are addressed in this document. Although last published in 1991, much of this guidance for estimating future year emissions is still valid. There have been updates to some of the information provided in the 1991 guidance (BEA projection phase-out and EGAS and MOBILE model revisions, etc.) and therefore States should review additional documentation concerning emissions projections in U.S. EPA (1998e).

States may find it useful to examine techniques that have been applied in other areas where control strategy planning has been performed. In the 1991 guidance document, examples of emission projection preparation are recorded in a form suitable for input to a grid-based photochemical model. In the simplest sense, this approach relied on developing a growth factor and a control factor for each major source category.

<u>Recommendations.</u> States should review guidance on emission projections issued by the U.S. EPA in 1991 and additional revisions to this document (U.S. EPA, 1998e). States should concentrate on making use of dominant source type information and variables historically shown to affect their economy. States should be aware of the uses of their emission projections, and factor relevant information into their estimates. States should review techniques previously used for emission projection efforts to utilize existing projections information. States should quality assure their emission projections using several methods designed to validate the spatial and temporal allocations, as well as any speciation that may be calculated. States should review past emission projection efforts as part of any subsequent review performed for the reasons identified in Section 5.0.

16.0 How Do I Assess Model Performance And Make Use Of Diagnostic Analyses?

Results of a model performance evaluation should be considered prior to using modeling to support an attainment demonstration. Performance of an air quality model can be evaluated in two ways: (1) how well is the model able to replicate observed concentrations of ozone and/or precursors, and (2) how accurate is the model in characterizing sensitivity of ozone to changes in emissions? The modeled attainment test recommended in Sections 3.1 and 3.2 uses models to predict sensitivity of predicted ozone to controls and then applies resulting relative reduction factors to *observed* (rather than modeled) ozone. Thus, while both types of performance test are important, the second type is the most important. Unfortunately, it is also more difficult to do.

Diagnostic analyses are potentially useful for several reasons. First, these analyses can be used to better understand why the air quality model predicts what it does. This yields further insight into whether or not the predictions are plausible. Second, diagnostic analyses provide information which helps prioritize efforts to improve/refine model inputs. Third, diagnostic tests can provide insight into which control strategies may be the most effective for meeting the ozone NAAQS. Fourth, diagnostic analyses can be used to assess how "robust" a control strategy is. That is, do I reach the same conclusion regarding adequacy of a strategy when using a variety of assumptions regarding current conditions?

In this section, we first identify methods which may be useful for evaluating model performance. We then discuss each of these methods in greater detail. We next note that there is no single method which offers a panacea for evaluating model performance. We recommend that performance be assessed by considering a variety of methods, much as is done in a weight of evidence determination. We then identify methods for performing diagnostic analyses. We conclude by identifying several potentially useful diagnostic tests which States should consider at various stages of the modeling analysis.

16.1 How Can I Evaluate Performance Of An Air Quality Model?

As noted above, model performance can be assessed in one of two broad ways: (1) how accurately does the model predict observed concentrations?, and (2) how accurately does the model predict *responses* of predicted air quality to changes in inputs? An example of the latter type of assessment is, "how accurately does the model predict relative reduction factors (RRF)?"

Given existing data bases, nearly all analyses have addressed the first type of performance evaluation. The underlying rationale is that if we are able to correctly characterize changes in concentrations accompanying a variety of meteorological conditions, this gives us some confidence that we can correctly characterize future concentrations under similar conditions. Computer graphics, ozone metrics, precursor metrics and observational models are all potentially useful for evaluating a model's ability to predict base case air quality.

The second kind of model performance assessment can be made in several ways. One

way is by looking at predicted differences on weekends vs. week days, provided reliable emissions estimates are available for both, and differences in weekend/week day emissions are substantial. A second way is to examine predicted and observed ratios of "indicator species". If observed ratios of indicator species are very high or very low, they provide a sense of whether further ozone production at the monitored location is likely to be limited by availability of NOx or VOC. Agreement between paired observed and predicted high (low) ratios suggests a model may correctly predict sensitivity of maximum (hourly) ozone at the monitored locations to emission control strategies. Thus, use of indicator species methods shows potential for evaluating model performance in a way which is most closely related to how models will be used in attainment demonstrations. We recommend that greater advantage be taken of these methods in the initial demonstration and in subsequent reviews. A third way for assessing a model's performance in predicting sensitivity of ozone to changes in emissions is to compare projections after the fact with observed trends.. One reason States should retain data files and output generated in simulating the control strategy selected for a SIP is to facilitate retrospective analyses. As explained in Section 5.0, these analyses provide potentially useful means for diagnosing why a strategy did or did not work as expected. They also provide an important opportunity to evaluate model performance in a way which is closely related to how models are used to support an attainment demonstration.

States can assess model performance using graphics, ozone metrics, predictions of precursor concentrations, corroborative analyses with observational models, weekend/weekday comparisons, ratios of indicator species and retrospective analyses with observed air quality and emission trends. These methods are described in the following subsections. For the 8-hour ozone NAAQS, States should compare 1-hour observations and predictions as well as observed and predicted 8-hour daily maxima.

16.1.1 How Can I Use Graphics To Make A "Big Picture" Assessment Of Model Performance?

States should refer to guidance in U.S EPA (1991) regarding use of graphics to evaluate model performance. Graphs plot predictions and observations. The 1991 guidance describes the following graphical displays: time series plots, tile plots, scatter plots and quantile-quantile (Q-Q) plots. Each of these graphics can also be used to display *differences* between predictions and their paired observations. Graphics are useful means for understanding *how* predictions and observations differ. For example, time series plots tell whether there is any particular time of day or day(s) of the week when the model performs poorly. Tile plots reveal geographic locations where the model performs poorly. Information from tile plots and time series may provide clues about where to focus quality assurance efforts for model inputs. Scatter plots and Q-Q plots show whether there is any part of the distribution of observations for which the model performs poorly. These plots are also useful for helping to interpret calculations of bias between observations and predictions. For example, they could show large differences between observations and predictions which just happen to balance, producing low estimated bias.

16.1.2 How Can Ozone Metrics Be Used To Assess Model Performance?

Ozone metrics produce numerical comparisons between observations and predictions. Appendix C in U.S. EPA (1991) identifies several metrics, as well as the mathematical formulae for calculating them. We recommend that comparisons of observations and predictions for 1-hour sampling times be used as well as comparisons of 8-hour daily maximum concentrations. One-hour comparisons of metrics provide a much larger data base for assessing model performance than would otherwise be available.

States should calculate metrics which are closely related to how model results are used in the recommended modeled attainment test (Sections 3.1 and 3.2). In the attainment test, models are used to calculate relative reduction factors (RRF) near monitoring sites. This is done by taking the ratio of the mean highest 8-hour daily maximum concentrations calculated for the future to that estimated with current emissions. Thus, the model's ability to predict observed mean 8-hour daily maxima is an important indicator of model performance. We recommend that States use the following set of ozone metrics to assess model performance during base periods corresponding to the selected episodes.[7]

1) Estimate bias between spatially paired means of observations and predictions of highest 8-hour daily maximum ozone concentrations. The "means" are the averages of 8-hour daily maxima observed and predicted over several days. Predicted values should be taken from grid cells "near" a monitor, as defined in Table 3.3. Predicted "nearby" 8-hour daily maximum for each day is calculated as illustrated in Figure 3.2. The comparison described in this test leads to a separate estimate of average bias in predicted 8-hour daily maximum ozone for each monitoring location.

2) Compute a correlation coefficient and display a scatter plot for the average observed and predicted 8-hour daily maxima used to estimate the average bias in test 1.

3) Compute a temporal correlation coefficient of observed and nearby predicted 8-hour daily maxima which are *spatially* averaged by day. If there are a sufficient number of monitoring sites for the analysis to be meaningful, it is also useful to group concentrations from those monitors that represent upwind, downwind and center city locations. Include time series and scatter plots of the results.

4) Prepare quantile-quantile plots of observed and predicted 8-hour daily maxima using (a) all data pairs (i.e., sample size = (# of stations)(# of days)), (b) spatially paired mean 8-hour daily maxima (i.e., sample size = # of stations), and (c) temporally paired spatially averaged 8-hour daily maxima (i.e., sample size = # of days).

---

[7] "Bias" and "fractional bias" are calculated as described in Appendix C to U.S. EPA (1991). In the text, we discuss 8-hour daily maxima. Similar tests could be performed for observations and predictions of 1-hour daily maxima.

5) Calculate the fractional bias for the pairings described in tests 4 (a), (b) and (c) above.

It is not possible to provide a definitive set of performance criteria for the preceding metrics, since we do not know how sensitive the model's response is to failure to replicate base case observations. However, we suggest the following as a performance goal for tests 1 and 5.

> The bias should be less than about 20% of the mean observed 8-hour daily maximum (test 1) and the spatially paired fractional bias (test 5(b)) should also be less than about 20%. These goals should be met at locations with monitored design values exceeding the NAAQS and at some of the other locations as well.

The "20%" goal is based on information like that shown in Figure 3-3. This figure suggests that mean RRF values do not appear to be very sensitive to predicted current 8-hour daily maxima unless these values are less than about 70 ppb. Seventy (70) ppb is about 20% less than 85 ppb, the lowest concentration which exceeds the level specified in the NAAQS. This goal may be a difficult one to meet, particularly if the mean predicted RRF is based on a single day's prediction. This is one reason why we suggest simulating several episodes in Section 12.0.

Performance goals for tests 2, 3 and 4 will, by necessity, have to be established on a case by case basis. This follows since the meaningfulness of any goal for correlation coefficients between paired observations and predictions depends on the variability in the observations. Assuming there is substantial variability in mean observed 8-hour daily maxima from site to site, a general goal is to have a moderate to high positive correlation between observations and predictions which is statistically significant.

We emphasize that the discussion in the preceding paragraphs addresses performance *goals* rather than *criteria*. Failure to meet one or more of the goals does not mean that an analysis cannot proceed. This decision can only be made on a case by case basis. However, success or failure of a model to meet performance goals should be noted, and may be considered in a weight of evidence determination.

The preceding five performance measures are oriented toward site by site comparisons. Other useful measures involve pooling these data to calculate overall bias and gross error for 1-hour predictions as well as for 8-hour daily maxima. The three most widely used pooled metrics for ozone have been unpaired 1-hour daily maximum concentrations, normalized bias and gross error. In past guidance we have identified performance criteria for these three measures. These criteria are based on results obtained in urban model applications (primarily in California) during the 1980's. This information may serve as one input in assessing how well a model performs.

If a State is primarily interested in showing that a strategy works for meeting the 8-hour NAAQS within or downwind of a nonattainment area, it may be useful to subdivide the monitoring sites into "downwind", "center city" and "upwind" categories on each modeled day rather than pool the entire data base. Pooled ozone metrics could then be calculated for each

category. Note that the identity of "upwind", "center city" and "downwind" sites could change from day to day. However, the aggregated metrics for all modeled days would be valid for the three defined *categories* of sites. This partitioning of sites may be considered using any ozone metric, providing there are enough sites available to support making such a partition. An output from such a grouping might be something like, "average bias or average fractional bias from all upwind sites, from all center city sites and from all downwind sites".

16.1.3 How Can I Use Available Precursor Observations To Evaluate Model Performance?

Ozone models have many degrees of freedom. This is another way of saying that you can predict similar ozone concentrations using a variety of combinations of (uncertain) inputs. Thus, a comparison of observed/predicted ozone is not a definitive assessment of model performance. Testing the ability of the model to predict other species, as well as ozone, is one means for increasing confidence in the results.

States should include an assessment of how well a model replicates observed VOC, VOC species treated explicitly in the model's chemical mechanism, CO, NOy, and NO_2 whenever the data base permits. One concern however about monitored precursor data is, what spatial scale do the monitored data represent? Monitored primary pollutant concentrations, like NO, CO, VOC and VOC species, could greatly depend on proximity of the monitor to one or a small number of sources. Models typically consider horizontal grid cells of 4 km or larger. Thus, the measurements may not be representative of the spatial scale the model is addressing, and the comparison between monitored and modeled data becomes difficult to interpret. This mismatch between models and monitored data is referred to as "incommensurability".

Potential problems introduced by incommensurability may be reduced in at least three ways. First, States should focus on secondary species like NO_2, or aggregates including secondary species, like NOx or NOy. The rationale for this strategy is that time is required for these secondary species to form, thus permitting greater mixing on the scales assumed in the model. A second strategy is to consider *ratios* of primary (or secondary) pollutants which tend to co-vary. For example, observed ratios of one or more selected VOC species to CO are likely to be less variable than concentrations of the individual pollutants. Therefore, the ratios may be more representative of the scales considered in the model. A third way for reducing incommensurability is to use metrics which entail spatial averaging in some manner. Comparisons between spatially averaged observations of VOC at 3 monitoring sites with spatially averaged model predictions "near" the 3 sites is an example of such an approach.

16.1.4 How Can I Use Corroborative Analyses With Observational Models To Help Evaluate Air Quality Model Performance?

Recently, techniques have been developed to embed procedures within the code of an air quality model which enable users to assess contributions of specific source categories or of

specific geographic regions to predicted ozone at specified sites (ENVIRON (1997), Yarwood, et al. (1997, 1997a), Morris, et al., (1997), Yang, et al., (1997, 1997a)). These source attribution procedures characterize what the air quality model *says* are the effects of targeted areas or sources on predicted air quality. Provided speciated VOC data are available at a site, source attributions estimated with these approaches can be compared with those obtained using other models which rely directly on observed air quality data.

The chemical mass balance model (Watson (1997)) is probably the most directly applicable observational approach for this purpose, since it can focus on the same day(s) considered with the air quality model. Cautions raised previously about representativeness of the monitored data continue to apply. Available multi variate statistical models (see, for example, Henry, et al., (1994) and Henry (1997, 1997a, 1997b)) may provide a more qualitative means for assessing an air quality model's performance. Multi variate statistical models work by examining temporal variability in monitored precursor species at a single site or spatial variability on one or a few occasions at many sites. A qualitative comparison is possible if one can contrast observations on days when winds suggest a source contribution is unlikely vs. days when a contribution is likely, or at locations where a source category is important vs. those where it isn't. If the observational approach suggests a major change in a source category contribution, and the air quality model also suggests that category is important or unimportant under similar wind conditions, the observational model lends credence to the air quality model's predictions.

16.1.5 What Data Bases Reflecting Changes In Emissions Are Available To Evaluate Model Performance?

Activity levels and patterns, leading to precursor emissions from mobile, area and some point sources, may differ on weekends vs. week days. If these differences are substantial, simulating weekend as well as week days could provide a means for evaluating how accurately a model predicts the effects of changing emissions.

Weekend/week day information could be used in one of two ways. The first way is to compare mean predicted and mean observed 8-hour daily maxima at each monitoring site for weekends vs. week days. If there are a sufficient number of monitors available, it is also desirable to make these comparisons for categories of monitors, grouped according to whether they represent "downwind", "center city" or "upwind" conditions. Tests 1-5, described in Section 16.1.2, as well as other tests, could be applied first for week days and then for weekend days. If the performance is adequate for both weekends and weekdays, this suggests that the model is accurately characterizing composite effects of different meteorological conditions and different emissions.

A second way for using weekend/week day information is to first screen the available data to identify weekend days and week days for which meteorological conditions are "similar". For example, for urban analyses, wind orientation, daily maximum surface temperature, presence of precipitation and maximum mixing height might be considered for this purpose. If similar sets

of meteorological conditions are identified for weekends and week days, *changes* in mean observed 8-hour daily maxima can be compared with *changes* in mean predicted 8-hour daily maxima for each monitor site, as well as for groups of sites characterized as "upwind", "center city" and "downwind". Tests like those described in Section 16.1.2 may be used for this purpose. If predicted changes generally provide an unbiased estimate of observed changes, this suggests that the model characterizes effects of changing emissions accurately.

We need to mention several caveats regarding weekend/weekday comparisons. First, changes in emissions between week days and weekends may be small compared to uncertainties associated with the weekend and week day estimates. Second, dividing a (already small) sample into weekends and week days may mean that conclusions are based on few comparisons. Third, identifying "similar" meteorological conditions may be somewhat subjective. Finally, the changes between weekend and week day emissions may be considerably less than the changes needed to meet the NAAQS in some areas. Since the relationship between precursor and ozone concentrations may be nonlinear, the weekend/weekday comparisons may not be definitive evaluations. Despite these reservations, weekend/week day comparisons provide one of a relatively few means for evaluating a model's ability to accurately predict changes in ozone concentrations. States should include these comparisons in their efforts to evaluate model performance, whenever feasible.

16.1.6 How Do I Use Ratios of Indicator Species To Evaluate Model Performance?

A performance evaluation which includes comparisons between modeled and observed ratios of indicator species carries with it a large potential advantage. Such a comparison may reveal whether the model is predicting sensitivity of ozone to VOC and/or NOx controls correctly. That is, when the model predicts ratios within a certain range, predicted ozone is sensitive to changes in a particular one of its precursors. Within another range of ratios, predictions are sensitive to changes in another precursor. If a model predicts *observed* ratios of indicator species such that observed and predicted ratios fall within the same range of ratios, this provides some confidence that the predicted *change* in ozone may be accurate.

For ozone modeling applications, uses for ratios of indicator species are described by Sillman (1995, 1997, 1998) and by Sillman, et al. (1997a). The authors of these references have shown several ratios of indicator species to be good indicators of whether peak predicted (i.e., *modeled*) ozone is likely to be most sensitive to reductions in VOC or NOx. Many of the species discussed require measurements beyond those which have been routinely made by most State agencies. Of the ratios discussed, the following involve compounds or mixtures most amenable to measurement by State agencies: O_3/NOy, O_3/NOz[8] and O_3/HNO_3. States should review the Sillman and Sillman, et al. references for further details about measurement requirements.

Strength of the indicator species approach for assessing model performance depends on

---

[8]Note: NOz = NOy - NOx

an assumption that the model is accurately characterizing the relationships between indicator species and ozone. The validity of this assumption can be more readily tested in smog chamber experiments than can absolute predictions of ozone. A second precaution is that there may be a range of observed ratios for which the preferred direction of control is not clear. When this occurs, agreement between predictions and observations does not necessarily imply that the model correctly predicts sensitivity of ozone to changes in precursors. Third, this method requires more measurements than are commonly made. In some cases, it may be difficult to achieve the required precision with routine monitoring. Finally, much of the work done to date with indicator species has focused on peak hourly concentrations of ozone. Applicability of the approach to 8-hour averaging times has not yet been extensively tested. Despite these precautions, the approach of comparing predicted and observed ratios of indicator species provides a means of assessing a model's ability to accurately characterize sensitivity of predicted ozone to changes in precursors. States should use the method to help evaluate model performance, whenever feasible.

16.1.7 Are Retrospective Analyses Useful For Evaluating Model Performance?

Retrospective analyses compare past model air quality projections with observed trends in air quality and estimated trends in emissions. The approach is a direct assessment of what we are most interested in---does the model accurately predict changes in air quality? However, it is not as straightforward as it seems. Often, input estimates and assumptions used in past studies are ambiguous and the emissions trends are qualitative. Also, the past studies generally assume constant meteorology, which does not happen. One of the purposes of the reporting requirements described in Section 6.0 is to make it possible for others to replicate modeled analyses at future dates.

In Section 5.0, we noted that a retrospective analysis is an important means for diagnosing why a NAAQS has or has not been attained. Such an analysis provides assurance that improved air quality results from changes in emissions rather than meteorology and/or can identify reasons why satisfactory progress is not being observed. Retrospective analyses will have an ancillary benefit of providing an additional means for evaluating model performance. In order to ensure some planning for subsequent retrospective analyses and to promote some uniformity in the methods used for these analyses, they are probably best performed as part of a subsequent review rather than as supporting evidence in the initial SIP revision.

16.1.8 All Of These Performance Tests Have Shortcomings, So What Do I Do?

There is no single definitive test for evaluating model performance. All tests have strengths and weaknesses. Credence given to model results is increased if a variety of tests is applied and the outcomes either support a conclusion that the model is working well or, at least, are ambiguous. Thus, one can think of a model performance evaluation as a "mini-weight of evidence analysis" focused on the issue of how much credence to give model results in an attainment demonstration. Table 16.1 summarizes the tests and their corresponding objectives or

goals described in this guidance.

Table 16.1. Summary Of Methods To Evaluate Performance Of Air Quality Models

Method	Test(s)	Goals/Objectives
Big Picture Assessment Using Graphics	**-tile plots of observations & predictions.** **-tile plots of *differences* in observations & predictions** **-scatterplots & Q-Q plots** **-time series plots**	**-complement ozone metrics** **-determine whether performance is worse for high vs. low observations** **-look at spatial patterns of performance--is performance better downwind than upwind?** **-Focus diagnostic tests on certain times/locations**
Ozone metrics	**-bias pred/obs mean 8-hr (& 1-hr) daily maxima near each monitor** **-fractional bias pred/obs mean 8-hr (& 1-hr) daily maxima near each monitor** **-correlation coefficients, all data, *temporally paired means*, spatially paired means** **-bias (8-hr daily max & 1-hr obs/pred), all monitors** **-gross error (8-hr daily max & 1-hr obs/pred), all monitors** **-partition pooled data base into "upwind", "center city" & "downwind" sites. Repeat analyses** **-scatterplots & Q-Q plots of 8-hr & 1-hr metrics**	**--20% most monitors (8-hr comparisons only)** **--20% most monitors (8-hr comparisons only)** **-moderate to large positive correlations** **--5-15%** **--30-35%** **-get a better idea of what parts of the distribution of predictions & obs agree or disagree & whether there is any obvious pattern to the model's performance**

Table 16.1. Summary Of Methods To Evaluate Performance Of Air Quality Models (continued)

Method	Test(s)	Goals/Objectives
Precursor concentrations	**Similar to ozone metrics. Focus on** **-secondary species (NO_2, NOy, NOx, NOz)** ***-ratios* of co-varying species (VOC, or VOC species/CO, VOC or VOC species/NOx)** **-spatially averaged predictions of the above or of primary species**	**-provide means for assessing whether model performance suffers in other respects if model is tuned to produce better ozone metrics** **-no specific goal for performance of model in predicting precursor concentrations is given.**
Observational models	**compare source attribution estimates with observational models** **-CMB** **-Multivariate models**	**-source attribution & CMB identify similar source categories as being important contributors to observed precursor concentrations** **-day to day variability in air quality model's source attribution & observations or mulit variate models is consistent** **-these are qualitative comparisons**

Table 16.1. Summary Of Methods To Evaluate Performance Of Air Quality Models (continued)

Method	Test(s)	Goals/Objectives
Weekend/week day comparisons	-compare previously identified ozone (& precursor) metrics on weekends vs. weekdays -if data base permits, partition data base into meteorlogical classes. For each class compare *differences* in weekday vs. weekend predictions with *differences* in weekday vs. weekend observations. -pool data base to compute bias and gross error on weekends and week days. -if data base permits, partition pooled data base into "upwind", "center city" and "downwind" bins & perform the previously identified pooled tests.	Objective is to test model's ability to accurately reproduce effects of *changing* emissions Same performance goals as mentioned for ozone metrics and precursor comparisons.
Ratios of indicator species	compare predicted and observed ratios at time of maximum observed ozone at each site. The following ratios are recommended for comparison of predictions & observations -O_3/NOy -O_3/NOz -O_3/HNO_3	Guidance refers to Sillman references to identify ratios where max.hourly ozone is likely limited by NOx & ratios where availability of VOC limits max.ozone. Predictions & observations should fall into the same class (i.e., VOC-limited cases, NOx-limited cases, cases where it is too close to call)

Table 16.1. Summary Of Methods To Evaluate Performance Of Air Quality Models (concluded)

Method	Test(s)	Goals/Objectives
Retrospective analyses	**project ozone to a future (preferably sooner than attainment date) year** **retain files** **update emission estimates at future year & note observed future ozone** **characterize future met.episodes & model in future**	**Recommended primarily for a subsequent review.** **note agreement/disagreements between projected ozone & subsequent observations** **perform diagnostic tests to determine whether disagreement is due to** **-differences in projected emissions vs. emissions estimated at a future date** **-differences in assumed meteorlogical conditions** **-a combination of different meteorological and emissions assumptions** **-one or more limitations in the model.**

Finally, we need to address the issue of adjusting model inputs to improve model performance. One of the reasons we recommend a variety of tests for model performance is to reduce the possibility of "getting the right answer for the wrong reason". We recognize however, that many of the inputs to models have associated (often unknown) uncertainties. It is acceptable to adjust inputs within reasonable bounds to improve performance, providing it does not result in poorer performance in any of the several measures of performance which we recommend in Sections 16.1.1 - 16.1.7. If such an adjustment is made, it should be documented and accompanied by an explanation as to why those implementing the protocol believe it is justified.

Recommendations. States should undertake a variety of performance tests. Results from a diverse set of tests should be documented and weighed to qualitatively assess model performance. Provided suitable data bases are available, greatest weight should be given to tests which assess model capabilities most closely related to how the model

is used in the modeled attainment test. A narrative describing overall assessment of model performance should be included among the material submitted to support a recommended SIP revision requiring a demonstration of attainment.

16.2 How Can I Make Good Use of Diagnostic Tests?

Diagnostic tests are performed using one of two broad approaches. The first of these consists of tests in which sensitivity of air quality predictions to perturbations in one or a combination of model inputs is examined. This is the more traditional of the two approaches and has a longer track record. When it is applied, States should recall how model outputs are used in the modeled attainment test recommended in Section 3.0. That is, models are used in a relative sense to provide relative reduction factors. Relative reduction factors are obtained by taking the ratio of mean 8-hour daily maximum concentrations obtained with future vs. current emissions.

The second type of diagnostic test is one in which means for "tracking" the importance of various phenomena contributing to predicted ozone at a location are embedded within the code of an air quality model. This generally increases running time and should not be used unless a vendor or someone very familiar with the computer code has installed this capability and performed benchmark tests to ensure that the model, with and without the code revisions, yields identical results. The major advantage of this latter type of diagnostic test is that it reduces the number of model simulations needed to obtain insights about what is causing high or unexpected ozone predictions. Thus, once the initial effort to develop this capability has been expended, a better understanding of why predictions are the way they are can be obtained relatively efficiently.

16.2.1 Use Of Sensitivity Tests

Outcomes of sensitivity tests are useful for several purposes. First, the tests can be used to assess the robustness of a control strategy. For example, States can consider effects of assumed boundary conditions and meteorological assumptions on predicted effectiveness of a control strategy. If the control strategy appears to work for a variety of assumptions, this increases confidence in the model results. Second, models used to support ozone NAAQS attainment demonstrations are resource intensive. Sensitivity tests provide a means for prioritizing use of resources in applying the model. For example, how sensitive are relative reduction factors to use of more vertical layers or smaller grid cells? Is using 4 km (rather than 12 km) grid cells more important than simulating many days? Third, sensitivity tests may help prioritize additional data gathering efforts so that a better subsequent review/diagnosis can be performed at the time of required attainment. Finally, sensitivity analyses could be useful for prioritizing control efforts or for noting sensitivity of predictions to uncertainties in the current or future emission inventory.

Sensitivity tests can and should be applied throughout the modeling process, not just

when model performance is being evaluated. Tests should be selected on a case by case basis by those implementing the modeling/analysis protocol. We present a sequence of activities likely to be followed in applying an air quality model. Under each activity we list some sensitivity tests which might be useful to resolve certain issues which may occur in some locations. The list is intended for illustrative purposes. The identified tests are not mandatory, nor is the list a comprehensive one.

Model Setup

-Change boundary conditions (is domain size adequate? do I need to consider using a nested regional model?).

-Alter initial conditions (do I need to extend the ramp-up period I have selected?).

Performance Evaluation/Troubleshooting

-Alter grid cell size and/or number of vertical layers considered (how are predicted base case 8-hour daily maxima affected?).

-Perturb specific inputs (e.g., mixing height, cloud cover, etc.) which might explain why certain processes are identified as important by process analysis (see Section 16.2.2) (are results affected by perturbations within reasonable bounds? what additional measurements should I try to make for a better subsequent review?).

Strategy Selection

-Simulate across-the-board reductions in VOC emissions, in NOx emissions and combinations of the two (what direction should I be thinking of for my control strategy?).

-Simulate reductions in point vs. area vs. mobile source emissions (what types of sources should my strategy focus on?).

-Simulate reductions in boundary conditions in concert with reduced emissions in the nonattainment area (will my strategy need additional help from regional controls in order to succeed?).

Uncertainty Estimates For Qualitative Use In Weight Of Evidence Determinations

-Simulate selected strategy starting with a different current inventory, reflecting reasonable uncertainties in current emissions.

-Simulate selected strategy, but with different (reasonable) growth projections.

-Perturb meteorological inputs, like mixing heights or cloud cover, which may be poorly characterized but which earlier analyses have suggested may be important in affecting base case predictions.

-Simulate selected strategy using different grid cell sizes and/or a different number of vertical layers.

16.2.2 Use of Process Analysis

Occasionally a review of a graphical display, like a tile diagram, may indicate a limited

number of locations or incidents which bear further investigation. Diagnostic tests may be used to perform focused analyses on these sites or incidents. These tests entail a more detailed look at a time series of predictions and (if available) observations at or above a site, including chemical species, winds and mixing. The examinations can be done qualitatively. However, more quantification is possible using the second type of diagnostic test described at the beginning of this subsection. A procedure called "process analysis" is an example of the second type of diagnostic test. Process analysis has been used to assess relative importance of various model assumptions as well as simulated physical and chemical phenomena contributing to a predicted ozone concentration at a particular time and location (Jeffries (1994, 1997), Jeffries, et al. (1996), Jang, et al. (1995), Lo, et al. (1997)).

Process analysis requires a substantial amount of expertise to be interpreted to full advantage. However, useful insights are also possible with less detailed analyses. The procedure focuses on selected grid cells. Process analysis then takes advantage of the fact that a numerical grid model addresses physical and chemical factors affecting ozone in a sequential manner. For example, a typical sequence followed in a model for each time step (e.g., 1 hour) might be (1) advection of ozone and precursors present at the beginning of the time step, (2) precursor emissions added during the time step, (3) vertical diffusion of the advected material and fresh emissions, (4) estimated cloud cover and its effects on photolysis rates, (5) atmospheric chemistry involving advected and diffused material with fresh emissions, and (6) deposition of certain compounds. Process analysis examines incremental effects on changes in ozone predictions from hour to hour attributable to each of the processes described above. In this way, one gets a sense of how important each process is as a contributor to predicted ozone at a specified time and location.

If a focused diagnostic analysis, such as one obtained with process analysis, suggests a particular model prediction may be an artifact of a model assumption rather than a result of real chemical/physical atmospheric processes, States may wish to go back to the meteorological or emissions model to verify that the inputs and assumptions that have been used are correct. If a prediction is the result of an apparent artifact which cannot be resolved, States may discount that prediction in the attainment demonstration.

Recommendations. States should include diagnostic analyses throughout the modeling process used to help select a control strategy which demonstrates attainment. These analyses should include sensitivity tests to assess robustness of a proposed strategy and consequences of simplifying assumptions made in the modeling. Additional sensitivity tests may be warranted on a case by case basis. Sensitivity of relative reduction factors to input perturbations should be a prime focus of the tests. Provided capabilities have been properly installed and tested, States may use versions of a model's code which contain capability for tracing importance different phenomena as contributors to predicted ozone concentrations at selected locations.

Table 16.2 shows examples of diagnostic tests which may be useful during different

stages of a modeling analysis.

Table 16.2. Potentially Useful Diagnostic Tests At Various Stages Of Modeling

Stage of Modeling	Test(s) (Examples)	Purpose(s)
Model Setup	-change boundary conditions	-is domain size sufficiently large? -do I need to use a nested regional model or will an urban scale model suffice?
Performance Evaluation & Troubleshooting	-alter specific (uncertain) inputs (e.g., mixing heights, cloud cover). -alter grid cell size or number of vertical layers considered.	-are ozone predictions improved? -what is the effect on other performance tests (e.g., precursor comparisons, weekend/weekday differences, indicator species ratios)? -what priorities should I assign to various kinds of improved measurements?
Strategy Selection	-simulate across the board reductions in VOC, NOx emissions & combinations of the two -simulate reductions in mobile vs. point vs. area source categories -perform the two preceding sets of tests with and without changing boundary conditions	-what sorts of strategies (VOC vs. NOx, emphasis on which source types) should I be considering? -will additional regional reductions in precursor emissions be necessary?

Table 16.2. Potentially Useful Diagnostic Tests At Various Stages Of Modeling (concluded)

Stage of Modeling	Test(s) (Examples)	Purpose(s)
Estimating Uncertainty	-simulate alternative base cases in emission estimates & project AQ from the alternative bases -simulate future AQ using alternative (reasonable) growth assumptions --includes different growth rates --different placement of new sources -perturb meteorological inputs which cannot be well characterized with available data simulate selected strategy using smaller (e.g., 4 km) grid cells	-assign a range (e.g., ± 1 std.dev.) of predicted RRF's based on distribution of predicted future & current mean 8-hour daily maxima near critical monitoring sites estimate range (e.g., ± 1 std.dev.) of future site specific design values with the range of RRF's. use the preceding information qualitatively as an input in a WOE determination
Focused performance analysis	-process analysis	-do suspicious looking results make physical sense?

17.0 References Cited In Part II

Alpine Geophysics, Inc., (1995), *The Emissions Modeling System (EMS-95) User's Guide*, Alpine Geophysics, Inc., Boulder CO.

Blanchard, C.L., F.W.Lurmann, P.M.Roth, H.E.Jeffries and M.Korc, (1999), "The Use of Ambient Data to Corroborate Analyses of Ozone Control Strategies", *Atmospheric Environment, 33*, pp. 369-381.

CARB, (1996), *Performance Evaluation of SAQM in Central California and Attainment Demonstration for the August 3-6, 1990 Ozone Episode.*

Causley, M.C., J.L. Fieber, M. Jimenez and L. Gardner, (1990), *User's Guide for the Urban Airshed Model, Volume IV: User's Manual for the Emissions Preprocessor System*, EPA-450/4-90-007D, U.S. EPA, Research Triangle Park, NC 27711, (NTIS No.: PB91-131250).

Chang, J.S., S.Jin, Y.Li, M.Beauharnois, C.L.Lu and H.Huang, (1997), *SARMAP Air Quality Model*, Prepared by Atmospheric Sciences Research Center, State University of New York (Albany).

Cox, W.M. and S.Chu, (1996), "Assessment of Interannual Ozone Variation in Urban Areas from a Climatological Perspective", *Atmospheric Environment 30*, pp.2615-2625.

Dennis, R.L., (1994), "Using the Regional Acid Deposition Model to Determine the Nitrogen Deposition Airshed of the Chesapeake Bay Watershed", *Ch.21, Atmospheric Deposition of Contaminants to the Great Lakes and Coastal Waters*, J.E. Baker (ed.), SETAC Press, Pensacola, FL.

Deuel, H.P. and S.G.Douglas, (1998), *Episode Selection for the Integrated Analysis of Ozone, Visibility and Acid Deposition for the Southern Appalachian Mountains*, Draft Technical Report Systems Applications International, Inc. (SYSAPP-98/07), prepared for the Southern Appalachians Mountains Initiative.

Dolwick, P.D., C.Jang and B.Timin, (1999), "Effects of Vertical Grid Resolution on Photochemical Grid Modeling", *Paper 99-309*, Presented at 99th AWMA Meeting, St.Louis, MO, (June 1999).

Douglas, S.G., R.C. Kessler and E.L. Carr, (1990), *User's Guide for the Urban Airshed Model, Volume III: User's Manual for the Diagnostic Wind Model*, EPA-450/4-90-007C, U.S. EPA, Research Triangle Park, NC 27711, (NTIS No.: PB91-131243).

ENVIRON International Corporation, (1997), *User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx)*, Novato, CA, April 17, 1997.

Geron, C., A.Guenther and T.Pierce, (1994), "An Improved Model for Estimating Emissions of Volatile Organic Compounds from Forests in the Eastern United States", *J.Geophys.Res. 99*, pp.12,773-12,791.

Grell, G.A., J.Dudhia and D.R. Stauffer, (1994), *A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5)*, NCAR/TN-398+STR, 138 pp.

Haney, J.L. and S.G. Douglas, (1996), *Analysis of the Effects of Grid Resolution on UAM-V Simulation Results for the Atlanta Nonattainment Area*, Systems Applications International Technical Memorandum SYSAPP-96/68, prepared for Southeast Modeling Center participants, October 21, 1996.

Henry, R.C., C.W.Lewis and J.F.Collins, (1994), "Vehicle-related Hydrocarbon Source Compositions from Ambient Data: The GRACE/SAFER Method", *Environmental Science and Technology 28*, pp.823-832.

Henry, R.C., (1997), "Receptor Model Applied to Patterns in Space (RMAPS) Part I. Model Description", *J.AWMA 47*, p.216.

Henry, R.C., (1997a), "Receptor Model Applied to Patterns in Space (RMAPS) Part II. Apportionment of Airborne Particulate Sulfur from Project MOHAVE", *J.AWMA 47*, p.220.

Henry, R.C., (1997b), "Receptor Modeling Applied to Patterns in Space (RMAPS) Part III. Apportionment of Airborne Particulate Sulfur in Western Washington State", *J.AWMA 47*, p.226.

Jang, J.C., H.E.Jeffries and S.Tonnesen, (1995), "Sensitivity of Ozone Model to Grid Resolution Part II: Detailed Process Analysis for Ozone Chemistry", *Atmospheric Environment 29* (21), p.3114.

Jeffries, H.E., (1994), "Process Analysis for UAM Episode 287", memo to Brock Nicholson, NC DEHNR, April 8, 1994. Available at the following Internet address ***"ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/projprpt/uammod"***, then download file "panel287.pdf".

Jeffries, H.E., T.Keating and Z.Wang, (1996), "Integrated Process Rate Analysis of the Impact of Nox Emission Height on UAM-modeled Peak Ozone Levels", Final Topical Report, Prepared for Gas Research Institute, Chicago, IL. Available at the following Internet address: ***"ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/reports/gri"***, then download file "uncgri.pdf".

Jeffries, H.E., (1997), "Use of Integrated Process Rate Analyses to Perform Source Attribution for Primary and Secondary Pollutants in Eulerian Air Quality Models", Presented at U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, June 16-18, 1997. Available at the following Internet websites, ***"ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/ipradocs"***, then download file "sourcewksp.pdf", and ***"http://www.epa.gov/ttn/faca/stissu.html"*** and then download file, "Source Att WS-Proc.analysis/mass track--Jeffries".

Kumar, N., M.T.Odman and A.G. Russell, (1994), "Multiscale Air Quality Modeling: Application to Southern California", *J.Geophysical Research 99*, pp.5385-5397.

Kumar, N. and A.G. Russell, (1996), "Multiscale Air Quality Modeling of the Northeastern United States", *Atmospheric Environment* 30, pp.1099-1116.

LADCO, (1999), Personal communication.

Lo, C.S. and H.E. Jeffries, (1997), "A Quantitative Technique for Assessing Pollutant Source Location and Process Composition in Photochemical Grid Models", Presented at annual AWMA Meeting, Toronto, Ontario (1997). Also available at the following Internet address, ***"ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/ipradocs"***, then download file "PCA_awma.pdf".

Lyons, W.A., C.J. Tremback and R.A. Pielke, (1995), "Applications of the Regional Atmospheric Modeling System (RAMS) to Provide Input to Photochemical Grid Models for the Lake Michigan Ozone Study (LMOS)", *J.Applied Meteorology 34*, 1762-1786.

MCNC, (1999), Internet website describing the MAQSIP air quality model. ***"http://www.iceis.mcnc.org/products/maqsip/"***

Milanchus, M.L., S.T. Rao, I.G. Zurbenko, (1998), "Evaluating the Effectiveness of Ozone Management Efforts in the Presence of Meteorological Variability", *J. Air and Waste Management 49*, pp.174-188.

Morris, R.E., G.M.Wilson, S.B.Shepard and K.Lee, (1997), *Ozone Source Apportionment Modeling Using the July 1991 OTAG Episode for the Northeast Corridor and Lake Michigan Regions*, (DRAFT REPORT), Prepared for Mr. Dan Weiss, Cinergy Corporation, Plainfield, IN.

NOAA, (1999), website, *"www.arl.noaa.gov"*.

Odman, M.T. and C.L. Ingram, (1996), *Multiscale Air Quality Simulation Platform (MAQSIP): Source Code Documentation and Validation*, MCNC Technical Report ENV-96TR002-v1.0, 83pp.

Pielke, R.A., et al., (1992), "A Comprehensive Meteorological Modeling System-RAMS", *Meteor.Atmos.Phys., 49* , pp.69-91.

Scire, J.S., R.J. Yamartino, G.R. Carmichael and Y.S. Chang, (1989), *CALGRID: A Mesoscale Photochemical Grid Model. Volume II: User's Guide*, Sigma Research Corp., Concord, MA.

Scire, J.S., F.R. Francoise, M.E.Fernau and R.J.Yamartino, (1998), *A User's Guide for the CALMET Meteorological Model (Version 5.0)*, EarthTech, Inc., Concord, MA.

Seaman, N.L., et al., (1995), "A Multi-Scale Four-Dimensional Data Assimilation System Applied to the San Joaquin Valley During SARMAP. Part I: Modeling Design and Basic Performance Characteristics", *J.Applied Meteorology 34*, pp.1739-1761.

Seaman, N.L. and D.R. Stauffer, (1996), *SARMAP Meteorological Model. Final Report to San Joaquin Valleywide Air Pollution Study Agency*, Technical Support Division, California Air Resources Board, Sacramento, CA., 174 pp.

Seaman, N.L., et al., (1996b), "Application of the MM5-FDDA Meteorological Model to the Southern California SCAQS-1997 Domain: Preliminary Test Using the SCAQS August 1987 Case", *Ninth Joint Conference on Applications of Air Pollution Meteorology*, American Meteorological Society, Atlanta, GA., January 28-February 2, 1996.

Seaman, N.L., (1997), "Use of Model-Generated Windfields to Estimate Potential Relative Mass Contributions from Different Locations", Prepared for *U.S. EPA Source Attribution Workshop*, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet Website, ***"http://www.epa.gov/ttn/faca/stissu.html"***, then download file "source attribution workshop materials".

Seigneur, C., P. Pai, J. Louis, P.Hopke and D.Grosjean, (1997), *Review of Air Quality Models for Particulate Matter*, Document Number CP015-97-1a , Prepared for the American Petroleum Institute.

Sillman, S., (1995), "The Use of NOy, H_2O_2, and HNO_3 as Indicators for O3-NOx-ROG Sensitivity in Urban Locations", *J.Geophys. Res. 100*, pp.14175-14188.

Sillman, S., (1997), "The Method of Photochemical Indicators as a Basis for Analyzing O3-NOx-ROG Sensitivity", NARSTO critical review paper, to be published in *Atmospheric Environment*

Sillman, S., D.He, C.Cardelino and R.E. Imhoff, (1997a), "The Use of Photochemical Indicators to Evaluate Ozone-NOx-Hydrocarbon Sensitivity: Case Studies from Atlanta, New York and Los Angeles", *J.Air and Waste Mgmt. Assoc.* , In press.

Sillman, S., (1998), *Evaluating the Relation Between Ozone, NO_x and Hydrocarbons: The Method of Photochemical Indicators*, EPA/600R-98/022.

Systems Applications International, (1996), *Users Guide to the Variable Grid Urban Airshed Model (UAMV)*, SYSAPP 96-95/27r, Documentation available for review on US EPA Internet website, "***www.epa.gov/scram001/t22.htm#uamv***".

Tesche, T.W. and D.E. McNally, (1993a), *Operational Evaluation of the SARMAP Meteorological Model (MM5) for Episode 1: 3-6 August 1990*, Prepared for the Valley Air Pollution Study Agency by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993b), *Operational Evaluation of the SARMAP Meteorological Model (MM5) for Episode 2: 27-29 July 1990*, Prepared for the Valley Air Pollution Study Agency by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993c), *Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 1: 26-28 June, 1991*, prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993d), *Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 2: 17-19 July, 1991*, prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993e), *Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 3: 25-26 August, 1991*, prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993f), *Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 4: 20-21 June, 1991*, prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.

U.S. EPA, (1991), *Guideline for Regulatory Application of the Urban Airshed Model*, EPA-450/4-91-013.

U.S. EPA, (1991a), *Procedures for Preparing Emissions Projections*, EPA-450/4-91-014.

U.S. EPA, (1993), *Volatile Organic Compound (VOC)/Particulate Matter (PM) Speciation Data System (SPECIATE), Version 1.5*, EPA/C-93-013.

U.S. EPA, (1993a), *User's Guide for the Urban Airshed Model, Volume IV: User's Manual for the Emissions Preprocessor System 2.0, Part A: Core FORTRAN System and Part B: Interface and Emission Display System*, EPA-450/4-90-007d(R), NTIS Accession No. PB93-122380.

U.S. EPA, (1994a), Office of Mobile Sources, *User's Guide for MOBILE5*, EPA-AA-AQAB-94-01.

U.S. EPA, (1996a), *Photochemical Air Monitoring System 1996 Data Analysis Results Report*, EPA-454/R-96-006.

U. S. EPA, (1997a), *EIIP Volume I, Introduction and Use of EIIP Guidance for Emission Inventory Development*, July 1997, EPA-454/R-97-004a.

U. S. EPA, (1997b), *Point Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004b.

U. S. EPA, (1997c), *Area Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004c.

U. S. EPA, (1997d), *Mobile Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004d.

U. S. EPA, (1997e), *Biogenics Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004e.

U. S. EPA, (1997f), *Quality Assurance Procedures*, July 1997, EPA-454/R-97-004f.

U. S. EPA, (1997g), *Data Management Procedures*, July 1997, EPA-454/R-97-004g.

U.S. EPA, (1998a), *EPA Third-Generation Air Quality Modeling System, Models-3 Volume 9b Users Manual*, EPA-600/R-98/069(b).

U.S. EPA, (1998d), *National Air Pollutant Emission Trends, Procedures Document 1900-1996*, EPA-454/R-98-008, see ***"http://www.epa.gov/ttn/chief/ei_datahtml#ETDP"***.

U.S. EPA, (1998e), Guidance for making emissions projections, in preparation.

U.S. EPA, (1998f), Guidance on temporal allocations, spatial allocations and chemical speciation, in preparation.

U.S. EPA, (1999b), Proposed revision to 40CFR, Part 51, Appendix W.

U.S. EPA, (1999c), *Emission Inventory Guidance For Implementation Of Ozone And Particulate Matter National Ambient Air Quality Standards (NAAQS) and Regional Haze Regulations*, EPA-454/R-99-006, (April 1999).

Watson, J.G., (1997), "Observational Models: The Use of Chemical Mass Balance Methods", Prepared for the U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet website, ***"http://www.epa.gov/ttn/faca/stissu.html"***, then download "source attribution workshop materials".

Yang, Y.J., J.G.Wilkinson and A.G.Russell, (1997), "Fast, Direct Sensitivity Analysis of Multidimensional Air Quality Models for Source Impact Quantification and Area-of-Influence Identification", Prepared for the U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet website, ***"http://www.epa.gov/ttn/faca/stissu.html"***, then download file "SAW-Direct Sensi Analys.--PowerP.--T.Russell".

Yang, Y.J., J.G.Wilkinson and A.G.Russell, (1997a), "Fast, Direct Sensitivity Analysis of Multidimensional Photochemical Models", *Environmental Science and Technology*, In press.

Yarwood, G. and R.Morris, (1997), "Description of the CAMx Source Attribution Algorithm", Prepared for U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet website, ***"http://www.epa.gov/ttn/faca/stissu.html"***, then download file "source attribution workshop materials".

Yarwood, G., G.Wilson, R.E.Morris, M.A.Yocke, (1997a), *User's Guide to the Ozone Tool: Ozone Source Apportionment Technology for UAM-IV*, Prepared for Mr. Thomas Chico, South Coast Air Quality Management District, Diamond Bar, CA, March 28, 1997.

Glossary

Areawide design value	The highest design value monitored within a nonattainment area. At a site with three years of complete data, a design value is the average 4th highest 8-hour daily maximum ozone concentration observed over a consecutive 3 year period.
Modeled attainment demonstration	A modeled attainment demonstration consists of two parts: an analysis estimating emission levels consistent with attainment of the NAAQS, and a list of measures that will lead to the desired emission levels once growth is accounted for. The first (analysis) part consists of a modeled attainment test. It may also include an additional screening analysis and a review of a diverse set of model outputs and emissions, air quality and meteorological data for consideration in a weight of evidence determination to assess whether attainment of the NAAQS is likely with the proposed control strategy.
Modeled attainment test	This test takes the ratio of mean predicted future and current 8-hour daily maximum ozone concentrations averaged over several days and multiplies this ratio times the site-specific monitored design value at each monitoring location. If the product is $\leq$ 84 ppb near all monitoring sites, the test is passed.
Modeling system	This is a group of models used to predict ambient ozone concentrations. The group includes an emissions model which converts countywide emission information into gridded speciated emissions which vary diurnally and reflect environmental conditions. It also includes a meteorological model which provides gridded meteorological outputs and an air chemistry/deposition model which takes information provided by the emissions and meteorological models and uses it to develop gridded predictions of hourly pollutant concentrations.

Relative reduction factor (RRF)	The ratio of predicted 8-hour daily maximum ozone averaged over several days near a monitoring site with future emissions to corresponding predictions obtained with current emissions.
Screening test	A screening test is used to ensure that a proposed control strategy will be effective in reducing ozone at locations without an air quality monitor so that attainment is shown throughout a nonattainment area. It consists of two parts. The first part is to examine predictions everywhere within and near the nonattainment area to identify locations having predictions which are consistently higher than any near a monitored location. The second part is to compute a relative reduction factor for each flagged location and multiply these factors times the areawide design value for the nonattainment area. If results are ≤84 ppb at all flagged locations, the test is passed.
Weight of evidence determination (WOE)	This is a set of diverse analyses used to judge whether attainment of the NAAQS is likely. The credibility of each analysis is assessed and an outcome consistent with an hypothesis that the NAAQS will be met is identified beforehand. If the set of outcomes, on balance, is consistent with attainment, then the WOE can be used to show attainment. A weight of evidence determination includes results from the modeled attainment test, the screening test, other model outputs and several recommended analyses of air quality, emissions and meteorological data.